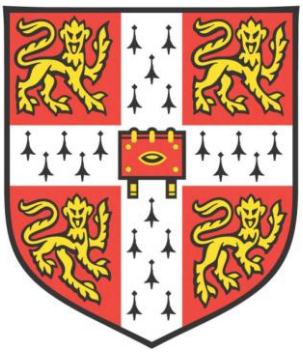
Characterising the role of TET2 in oestrogen receptor positive breast cancer



Rebecca Elizabeth Broome

Jesus College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. In accordance with the guidelines of the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine, this thesis is does not exceed 60,000 words.

Rebecca Elizabeth Broome

Signed:_____

Date:_____

Acknowledgements

Firstly, thank you to my supervisor, Professor Jason Carroll. I am hugely grateful for all the support, encouragement and guidance, and the opportunity to work in such an inspiring field of research with a wonderful team of people.

A special thank you to Dr Rasmus Siersbæk for the invaluable scientific discussions and experimental advice, patience, support and encouragement, without which this work would not have been possible. It was a real pleasure to work together on this project.

To the Carroll lab members past and present, notably Silvia, Sankari, Eva, Anca, Shalini, Sanjeev, Al, Sunny, Emily, Danya, Steve, Joe, Stacey, Kelly, Caryn, and Jill, thank you for all your help, and for making the lab such a fun place to work.

Special thanks to colleagues in bioinformatics, most notably Dr Igor Chernukhin and Dr Kamal Kishore, for all the useful discussions and invaluable expertise. Thank you also to Dr Shiqing Mao for the experimental support and scientific input, and to Dr David Tannahill for the advice and guidance.

I wish to thank the CRUK-CI core facilities, in particular the Proteomics team. I am also grateful to Genomics, the Genome Editing core, and Research Instrumentation & Cell Services for help in this work. I further acknowledge Cancer Research UK for funding these studies, and thank the Scientific Administration team for all their help along the way.

Finally, thank you to my friends and family, Mum and Derek, Dad, Laura, and Joe, and especially to Steve, for being a constant source of love, support and encouragement.

Summary

Characterising the role of TET2 in oestrogen receptor positive breast cancer

Rebecca Elizabeth Broome

The activity of the transcription factor oestrogen receptor α (ER) drives tumour development and metastasis in ER positive (ER+) breast cancer. This subtype constitutes up to three-quarters of breast cancer cases, and treatments directly targeting ER are currently the standard of care. ER operates as part of a signalling complex, co-ordinating with several other proteins to regulate gene expression. GATA3 is a transcription factor that has been closely linked to ER function, although its precise contribution to ER biology is not fully understood. The initial aim of this thesis was to further investigate the role of GATA3 in ER signalling. In doing so, GATA3 was shown to regulate the participation of the dioxygenase enzyme TET2 in the ER complex. The remainder of the thesis investigates the contribution of TET2 to ER biology.

Quantitative multiplexed rapid immunoprecipitation mass spectrometry of endogenous proteins (qPLEX-RIME) was used to assess changes to the ER complex in response to GATA3 knockdown. Upon GATA3 silencing, and its loss from the ER complex, TET2 stood out as the only additional protein that was significantly depleted as an ER interactor, whilst several proteins were significantly enriched in the ER complex, including the transcription factors LHX4 and ZBTB34. LHX4 and ZBTB34 are thought to be capable of binding methylated DNA, whilst TET2 is also implicated in DNA methylation pathways through its iterative conversion of methylated DNA to several additional DNA modifications, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. These results implied a potential role for GATA3 in modulating reading and writing of DNA modifications as part of the ER complex. In the context of the recently developing focus on TET2 in transcriptional regulation, the association between ER and TET2 was investigated further.

As few studies have successfully mapped endogenous TET2-chromatin interactions using chromatin immunoprecipitation (ChIP), the performance of various TET2 antibodies was evaluated using both ChIP and non-quantitative RIME. A commerical antibody was identifed that performed robustly in both techniques, and TET2 binding events were subsequently shown to constitute a near-total subset of ER binding events in both ER+ breast cancer cell lines and patient-derived xenograft models. ER was shown to bind at

several sites upstream of the TET2 promoter, supporting multiple existing studies proposing TET2 as an ER target gene. A key role for TET2 in the ER complex was further reinforced using non-quantitative RIME, where it was shown that TET2 associates not only with ER and GATA3, but additionally with several other key ER-associated proteins such as FOXA1, GREB1 and RARα. It was hypothesised that TET2, as an ER-regulated component of the ER complex, may help to sustain ER-regulated transcription in breast cancer.

Consistent with a role for TET2 in mediating the proper activity of the ER complex, loss of TET2 correlated with a depletion of global ER binding. This was associated with dysregulated expression of both ER and GATA3 target genes. The genes most markedly regulated by loss of TET2 in ER+ breast cancer cells included those related to cell cycle control, and providing functional evidence of this, growth of ER+ MCF7 cells was inhibited by TET2 knockdown. To investigate the TET2-ER relationship in more detail, the role of TET2 in regulating DNA modifications in ER+ breast cancer cells was examined. TET2 knockdown did not appear to induce changes in global DNA methylation, and whilst preliminary data suggested TET2 may modulate DNA methylation levels specifically at ER binding regions, this remains to be fully determined. However, additional studies showed that TET2 robustly regulates oxidation of methylated DNA to 5-hydroxymethylcytosine in ER+ breast cancer cells, indicating a role for this enzyme in the production and maintenance of 5-hydroxymethylcytosine at ER sites. The conclusion to this thesis puts the work presented into perspective and provides an outlook for future studies.

Contents

List of Tables

List of Figures

Key Abbreviations and Acronyms

AD1	Activation domain 1
AD2	Activation domain 2
AF-1	Activation function 1
AF-2	Activation function 2
AI	Aromatase inhibitor
AIB1	Amplified in breast cancer 1
ATP	Adenosine triphosphate
bp	Base pairs
BRCA1/2	Breast cancer gene 1/2
BRG1	Brahma-Related Gene 1
^{o}C	Degrees Celsius
CARM1	Coactivator-associated arginine methyltransferase 1
CBP	CREB-binding protein
cDNA	Complementary deoxyribonucleic acid
CGI	CpG island
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by high-throughput sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
CRUK-CI	Cancer Research UK Cambridge Institute
Da	Daltons
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBD	DNA binding domain
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid

DNMT	DNA methyltransferase
DSBH	Double-stranded beta helix
DSG	Discuccinimidyl glutarate
E2	17β-oestradiol
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ER	Oestrogen receptor
ERE	Oestrogen response element
FBS	Foetal bovine serum
FCS	Foetal calf serum
FDR	False discovery rate
FOXA1	Forkhead box A1
GATA3	GATA binding protein 3
GREB1	Growth regulation by estrogen in breast cancer 1
GREB1L	Growth regulation by estrogen in breast cancer 1-like protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HER2	Human epidermal growth factor receptor 2
Hi-C	High-throughput chromosome conformation capture
HMT	Histone methyltransferase
IgG	Immunoglobulin G
kb	Kilobase
kDa	Kilodaltons
LBD	Ligand binding domain
LC	Liquid chromatography
LHX4	LIM/homeobox 4
M	Molar

MACS	Model-based analysis for ChIP sequencing
MBD	Methyl binding domain
mESC	Mouse embryonic stem cells
MMS	Methyl midi-seq
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NCOA	Nuclear receptor co-activator
NCOR	Nuclear receptor co-repressor
OGT	O-linked N-acetylglucosamine (GlcNAc) transferase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDX	Patient-derived xenograft
PGR	Progesterone receptor
PRMT	Protein arginine methyltransferase
qPCR	Quantitative polymerase chain reaction
qPLEX-RIME	Quantitative multiplexed rapid immunoprecipitation mass spectrometry of endogenous proteins
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RARA/α	Retinoic acid receptor alpha
RFS	Relapse-free survival
RIME	Rapid immunoprecipitation mass spectrometry of endogenous proteins
RIPA	Radioimmunoprecipitation assay buffer
RPMI	Roswell Park Memorial Institute
RRHP	Reduced representation hydroxymethylation profiling
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SERD	Selective oestrogen receptor degrader

SERM	Selective oestrogen receptor modulator
siRNA	Small interfering ribonucleic acid
TBS	Tris-buffered saline
TDG	Thymine DNA glycosylase
TE	Tris ethylene diamine tetraacetic acid
TEAB	Tetraethylammonium bromide
TET	Ten-eleven translocation
TFF1/pS2	Trefoil factor 1
TMT	Tandem mass tag
TSS	Transcription start site
U	Units
UBC	Ubiquitin C
UHRF	Ubiquitin-like PHD and RING finger domain-containing protein
V	Volts
XBP1	X-box binding protein 1
ZBTB34	Zinc finger and BTB domain containing protein 34
3C	Chromosome conformation capture
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-methylcytosine
5mC	5-hydroxymethylcytosine

Chapter 1

Introduction

1.1 Breast cancer

Breast cancer accounts for one in four female cancers, with 1.67 million new cases diagnosed globally in 2012. It is the leading cause of cancer-related death amongst women in less developed regions of the world, and in more developed regions, is second only to lung cancer (Ferlay et al. 2015). Breast cancer can also occur in men, although this constitutes less than 1% of all cases (Korde et al. 2010). In the UK, ~55,000 new cases are diagnosed yearly (2014-2016 statistics, Cancer Research UK). Risk factors for breast cancer include both environmental and hereditary components. Mutations in the *BRCA1* and *BRCA2* genes are known to significantly increase the probability of developing breast cancer, and are associated with 5% of all cases (Ripperger et al. 2009). Other factors contributing to breast cancer risk include age and hormone exposure. In women, the latter can be linked to the number of full-term pregnancies a woman undergoes and the age at which full-term pregnancy first occurs, as well as the use of hormone replacement therapy (Key et al. 2001).

1.1.1 Breast cancer classification

Breast cancer is a heterogeneous disease that can be grouped into several different subtypes. Classifications are based on both the cell type of origin and the presence or absence of several key molecular markers. Tumours can arise from either the ductal or lobular epithelial tissue of the breast, with ductal malignancies being the most common (Li 2003). A hormonal component to breast cancer has long been established, since the demonstration by Beatson that removal of the ovaries, the major sex hormone-producing tissue in females, reduced tumour burden in a patient suffering with breast cancer

(Beatson 1896). As such, the molecular classification of breast cancers is largely based on the presence or absence of the hormone receptor oestrogen receptor α (ERα) (Sorlie et al. 2001). ERα negative tumours constitute three main subtypes, characterised either by high expression of keratins and laminin (basal-like), elevated expression of human epidermal growth factor receptor 2 (HER2+), or the broad expression of adipose and other non-epithelial cell genes (normal breast-like). In contrast, ERα positive tumours recapitulate many characteristics of the luminal epithelial cells of the breast, including expression of the transcription factors GATA binding protein 3 (GATA3), X-box binding protein 1 (XBP1), and hepatocyte nuclear factor 3α, also known as Forkhead box protein A1 (FOXA1) (Sorlie et al. 2001; Perou et al. 2000). Of the ERα positive and ERα negative classes, ERα positive tumours are by far the most common, accounting for over two-thirds of all breast cancer cases.

Molecular classification of breast cancer can help predict both risk of recurrence and treatment response. For example, the PAM50 classification is a gene expression assay which measures differential expression of 50 key genes to accurately classify breast tumours according to the subtypes described above (Parker et al. 2009; Ellis et al. 2011), and can be used to predict the benefit of hormonal therapy and chemotherapy across all breast cancer subtypes. More recently, genome-wide analysis of large patient datasets has facilitated further subclassifcation of breast cancers into 11 different integrative (IntClust) subtypes. This classification is based on a combination of genomic copy-number alterations and gene expression. The integrative clusters each have different prognostic values and may provide more accurate predictions of disease recurrence, facilitating more effective treatment and monitoring (Curtis et al. 2012; Ali et al. 2014; Rueda et al. 2019).

1.2 The oestrogen receptor (ER)

ER is a nuclear hormone receptor that acts as a ligand-dependent transcription factor. In the canonical mechanism of ER signalling, ER binds to its ligand, most commonly 17β-oestradiol (E2), a metabolite of oestrogen. This induces conformational changes leading to receptor homodimerisation, which facilitates ER binding at oestrogen response elements (EREs) where it regulates gene expression (Kumar et al. 1987). There are two subtypes of the oestrogen receptor, ERα and ERβ, encoded by different genes. These

proteins share some structural homology (Kuiper et al. 1996) and are both expressed in hormone-responsive tissues such as the ovaries, prostate, testis, and breast. However, ERα and ERβ display differing cellular distributions in the mammary gland (Speirs et al. 2002), indicative of their potentially divergent functions. ERβ has been proposed to have a growth suppressive rather than growth stimulatory role in the breast, and may antagonise the effects of ERα in both normal development and cancer (Speirs et al. 2002). Despite the potential of a role for ERβ in breast cancer, of the two subtypes, ERα (hereafter ER) is far more directly implicated in this disease, and has formed the basis of the majority of investigations into breast cancer aetiology and treatment.

1.2.1 ER in normal physiology and cancer

The activity of ER in both the normal breast and in cancer is driven primarily by oestrogens, a group of female sex hormones that play a fundamental role in mammary gland development. There are three major naturally-occurring oestrogens in women: oestrone, oestradiol, and oestriol. An additional oestrogen, oestetrol, is produced only during pregnancy. Oestradiol is the major circulating oestrogen during reproductive years, and is also the most potent of all four oestrogens (Baker 2013). As all these hormones act through ER, these four forms will henceforth be collectively referred to as oestrogen. Oestrogen drives ductal elongation, branching and lactational differentiation of the mammary gland after puberty, and is subsequently associated with epithelial proliferation in the breast during the menstrual cycle and in pregnancy (Brisken & O'Malley 2010). Mice possessing a mutated, oestrogen-unresponsive ER are developmentally viable but suffer reproductive defects (Lubahn et al. 1993), and selective ablation of ER in the mammary gland arrests its development at the prepubertal stage (Feng et al. 2007). ER acts in synchrony with several other hormones such as progesterone and prolactin to regulate these diverse developmental processes in a timely and co-ordinated manner (Brisken & Ataca 2015). However, whilst in the normal breast, ER expression is restricted to a subset (~7%) of cells (Shoker et al. 1999), during malignancy its activities in regulating cell proliferation are co-opted by cancer cells. As such, the majority of breast tumours exhibit ER expression as a key feature. In breast cancer, ER continues to regulate cell growth by driving expression of genes related to the cell cycle and proliferation, however these processes gradually escape normal control mechanisms, and oestrogen becomes a driver for tumour development (Carroll 2016). ER positive (ER+) breast cancers tend to have a

better prognosis than ER negative (ER-) subtypes, partly due to the fact that ER presents a tractable target for therapeutic intervention (Parker et al. 2009). However, even within the ER+ class, tumours possess marked heterogeneity and differ in their treatment response (Lacroix & Leclercq 2004; Curtis et al. 2012).

1.2.2 ER structure

ER is a 66 kDa protein, with a structure similar to other members of the nuclear receptor superfamily, such as the progesterone receptor, the androgen receptor, and the glucocorticoid receptor (Huang et al. 2010). ER possesses three major domains: an N-terminal domain, a DNA-binding domain (DBD), and a ligand-binding domain (LBD). Within this, the protein is divided into six functional domains, classified as A-F (Figure 1.1). Functional domains A and B make up Activation Function domain 1 (AF-1). This region acts in a ligand-independent manner and is important for ER dimerisation prior to DNA binding. Functional domain C constitutes the DBD, where ER-DNA interactions are mediated through two zinc finger motifs (Klinge 2001). Domain D contains a nuclear localisation signal and is the hinge region of the protein. Domains E and F contain both the LBD, and a second Activation Function domain, AF-2. In the LBD, 12 alpha helices (H1-H12) create a structure that facilitates the docking of oestrogen (Brzozowski et al. 1997), whilst AF-2 possesses both a nuclear localisation signal and a further dimerisation domain. Importantly, both AF-1 and AF-2 are key mediators of the interactions between ER and its cofactors, proteins recruited by the receptor that help it to achieve its transcriptional functions. AF-1 and AF-2 can act either independently or synergistically to recruit cofactors, with the hinge region (domain C) shown to aid this synergistic recruitment (Zwart et al. 2010).

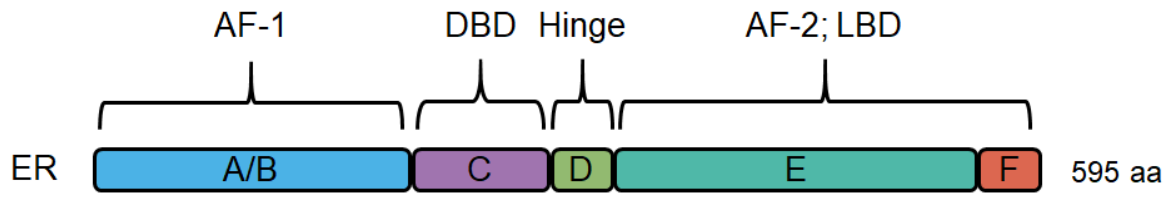


Figure 1.1. Structure of ER.

The N-terminus of ER (functional domains A/B) comprises activation function domain 1 (AF-1). Functional domain C contains the DNA binding domain (DBD), where two zinc finger motifs facilitate ER-DNA interactions. Domain D is the hinge region, whilst domains E and F comprise activation function domain 2 (AF-2), and the ligand binding domain (LBD), where endogenous ER ligands bind.

1.2.3 Oestrogen response elements (EREs)

Where present in the cytoplasm, ER is capable of indirectly affecting gene expression through non-genomic routes, for example through interacting with tyrosine kinases that co-ordinate intracellular signalling cascades (Madak-Erdogan et al. 2008). However, the majority of ER-regulated gene expression occurs through its ligand-activated assembly of transcription-regulating complexes at specific genomic regions. ER DNA binding motifs are known as oestrogen response elements (EREs). The canonical ERE is a 13-bp palindromic motif with the consensus sequence GGTCAnnnTGACC (Klinge 2001), and ER binds as a homodimer to this sequence. *In vitro* studies have showed that ER binding is preserved in the presence of minor variations in the canonical ERE (Schwabe et al. 1995), and chromatin immunoprecipitation followed by sequencing (ChIP-seq) to map ER-chromatin interactions has demonstrated that in many cases, only half this sequence is required for ER-DNA interactions (Joseph et al. 2010; Lin et al. 2007; Bourdeau et al. 2004). Whilst variations within the ERE can affect ER binding affinity, differences in the sequence of flanking regions can also influence the strength of ER binding (Anolik et al. 1995). Conferring further diversity on the mechanisms and affinity of ER binding, ER can also tether to DNA via protein-protein interactions with other transcription factors (discussed below) (Kushner et al. 2000; Safe 2001; Lin et al. 2007), which in some cases may mitigate the requirement of an ERE for effective oestrogen-mediated gene regulation.

1.2.4 ER acts mainly from enhancers

Transcription factors exert their effects on gene expression through interacting with different regulatory elements within the genome, including promoters, enhancers, insulators and boundary elements (Fulton et al. 2009; Palstra & Grosveld 2012). ChIP-seq has shown that ER predominantly associates with distal regulatory elements, and only ~3-5% of ER binding events occur in promoter-proximal regions (within 1-5 kb of the transcription start site, TSS) (Carroll et al. 2005; Carroll et al. 2006; Lin et al. 2007). Although the function of these promoter-distal ER binding sites was initially unknown, subsequent experiments confirmed these as enhancers, genomic regions where regulatory factors assemble to modulate gene expression at various distances from their target promoter (or promoters) (Pennacchio et al. 2013). Therefore, whilst a number of ER target genes (including *ESR1* itself) do possess promoter-proximal ER binding sites, most ER regulatory sites are distal to the genes they control (Carroll et al. 2006; Laganiere et al. 2005). Whilst ER binds primarily at enhancers, transcriptional regulation is effected ultimately at promoters, where the basal transcription machinery assembles to initiate mRNA assembly. It appears that the influence of ER function on promoter activity is mediated through chromatin looping that connects transcriptional regulatory complexes assembled at enhancers with their target promoters, translating the effects of ER at distal sites to TSSs (Fullwood et al. 2009; Kadauke & Blobel 2009; Pan et al. 2008).

Although different studies report varying numbers of ER binding events (even within the same cell line), it is generally concluded that there are many more ER binding sites in the genome than there are differentially regulated genes. This is partly consistent with the enhancer model, whereby multiple enhancers may regulate a single gene (Karnuta & Scacheri 2018). However, this could also imply that some of these ER sites might not have a transcriptional function in the cell types in which they are observed, and/or under the experimental conditions in which they are detected. Some of these apparently transcriptionally non-functional ER sites may exert transcriptional influences in different cell types or under different conditions (Carroll et al. 2006). Overall, enhancers appear to be poorly conserved, both across different species, and different tissue types (Pennacchio et al. 2013), suggestive of a role in regulating cell type-specific expression (Heintzman et al. 2009; Heinz et al. 2015). Taken together, these factors render it difficult to link enhancers directly to their target genes.

Several techniques have facilitated mapping of specific ER-regulated enhancer-promoter interactions, with the rationale that physical association between specific chromatin domains may provide evidence of a functional connection between these regions. Chromosome conformation capture (3C) (Dekker et al. 2002) has been used to identify, in a targeted manner, ER enhancers physically associated with the promoters of several key oestrogen-regulated genes, including *TFF1* (trefoil factor 1), *PGR* (progesterone receptor), and *GREB1* (growth regulation by estrogen in breast cancer 1) (Bonéy-Montoya et al. 2010; Deschênes et al. 2007; Pan et al. 2008). The subsequent development of a genome-wide approach to mapping factor-specific enhancer-promoter interactions, ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) allowed Fullwood et al. (2009) to probe ER-mediated chromatin looping in MCF7 cells. This demonstrated that the majority (86%) of ER-mediated chromatin loops anchoring two regulatory sites occur in the range of 10-100 kb. However, it was shown that overall, ER-mediated loops tend to connect more than two regulatory sites at once, highlighting the potentially diverse and complex nature of ER-regulated gene expression. More recently, Hi-C (high-throughput chromosome conformation capture), an unbiased technique for examining overall chromatin architecture, has been used to show that oestrogen induces a global reorganisation of chromatin structure in MCF7 cells (Mourad et al. 2014). Specifically, this showed that gene-rich chromosomes and areas of open and highly transcribed chromatin are brought into closer proximity by oestrogen treatment in these cells, presumably triggering interplay between the transcriptional regulatory machinery assembled at enhancers, and the promoters of target genes. Accordingly, analysis of these differentially-regulated loci showed an enrichment of oestrogen-induced genes and those related to proliferative control.

An additional phenomenon that may be useful in connecting distal enhancers to regulation of potential target promoters is the expression of enhancer RNAs (eRNAs). These are non-coding RNAs transcribed as a result of RNA polymerase II activity at enhancers. Oestrogen-induced eRNAs have been suggested to be important in ER-mediated gene expression through interactions with promoters, and may themselves have a role in oestrogen-induced chromatin looping (Li et al. 2013; Hah et al. 2013), although this is a contentious issue requiring further research. More recent attempts to infer functional links between specific *cis*-regulatory elements and their regulated genes have utilised CRISPR-based approaches, systematically disrupting transcription factor binding sites in

combination with studying the effect on the expression of nearby genes (for example Fei et al. 2019). The approaches described above, when used in combination, may can help narrow down those cell type-specific ER binding events that display functionality with respect to gene expression.

1.3 The ER complex

ER does not function in isolation, but co-operates with numerous proteins to achieve its transcriptional functions. Some of these proteins, known as cofactors, help the receptor to access its target sites, and some function as adapters mediating further protein-protein interactions. Additionally, others possess intrinsic enzymatic activity and can faciliate, for example, modification of histones and chromatin remodelling. Several of these key processes are described in the next sections.

1.3.1 ER-associated transcription factors

1.3.1.1 Pioneer factors

Although ER mediates its transcriptional actions predominantly through binding to DNA, it cannot achieve independent access to compacted chromatin. Pioneer factors are proteins that can associate with condensed chromatin and enable ATP-independent remodelling to facilitate the binding of further transcription factors (Zaret & Carroll 2011). The first pioneer factor characterised in relation to ER was FOXA1 (Carroll et al. 2005) which achieves its remodelling functions partly through its ability to displace linker histones (Cirillo et al. 1998; Cirillo et al. 2002). FOXA1 expression is considered a key marker of ER+ luminal tumours (Perou et al. 2000; Sorlie et al. 2001). ChIP-seq has shown that ~50% of ER binding events are co-occupied by FOXA1 in ER+ breast cancer cell lines (Hurtado et al. 2011), yet FOXA1 silencing results in decreased ER-chromatin occupancy even at sites where FOXA1 is not initially detected, suggesting it is capable of stabilising ER binding in both a direct and indirect manner. Depletion of FOXA1 in ER+ breast cancer cell lines has profound effects on oestrogen-regulated gene expression and results in impaired proliferation (Hurtado et al. 2011; Laganiere et al. 2005), indicative of its key role in ER-mediated transcription. Genomic binding of FOXA1 is not dependent on oestrogen treatment, suggesting that this factor demarcates regulatory sites prior to ER recruitment

to DNA, and hence determines the ER pathways that are regulated upon ligand stimulation (Glont et al. 2019; Hurtado et al. 2011). In line with this, differing distributions of ER binding are observed between primary tumours with good versus poor outcomes, and this is thought to be mediated by FOXA1 (Ross-Innes et al. 2012). Why FOXA1 behaves differently in terms of its binding distribution in cancers with different outcomes remains under investigation.

FOXA1 is not the only protein that provides pioneer functions for ER. The transcription factor transducin-like enhancer of split 1 (TLE1), shown to interact with histones but previously thought to mediate chromatin condensation (Palaparti et al. 1997; Sekiya & Zaret 2007), is important for ER binding at a subset of sites in ER+ breast cancer cells. Loss of TLE1 alters recruitment of RNA polymerase II at ER sites, associated with impared proliferation (Holmes et al. 2012). Other proteins possessing pioneer factor-like activity in relation to ER include pre-B-cell leukemia transcription factor 1 (PBX1) and activating enhancer-binding protein 2 gamma (AP2-γ), both of which also co-occupy ~50% of ER binding sites in ER+ breast cancer cell lines and are important for oestrogen-regulated gene expression and cell growth (Magnani et al. 2011, 2015; Tan et al. 2011). Analysis of FOXA1, PBX1 and AP2-γ binding events in the MCF7 genome showed that despite their potentially similar roles, these three proteins co-associate at ~30% of ER binding events, implying that in some instances they may act co-operatively in mediating ER access to chromatin (Jozwik & Carroll 2012).

1.3.1.2 GATA3

Another transcription factor that is consistently implicated in ER+ breast cancer is GATA3. Along with FOXA1, GATA3 is one of the minimal gene markers of ER+ tumours (Perou et al. 2000; Sorlie et al. 2001), and, similarly to FOXA1 motifs, GATA3 motifs are enriched around ER binding sites in MCF7 cells, implying a functional interplay between these three proteins (Carroll et al. 2005; Lin et al. 2007; Serandour et al. 2013). ER and GATA3 overlap at ~45% of ER binding events in E2-stimulated MCF7 cells, and co-expression of ER, GATA3 and FOXA1 together was shown to be required to confer oestrogen-responsive growth in an ER negative, oestrogen-unresponsive cell line (Kong et al. 2011). Eeckhoute et al. (2007) also showed a key role for GATA3 in oestrogen-mediated growth, demonstrating that GATA3 knockdown in ER+ T47D cells reduces their proliferative

response to oestrogen stimulation. Much like ER, GATA3 plays a fundamental role in mammary gland development, and conditional deletion of GATA3 results in severe maturation defects of the mouse mammary ductal tree during puberty (Kouros-Mehr et al. 2006). In breast cancer, loss of GATA3 is associated with a higher histologic grade, poor differentiation, and positive lymph nodes, concurrent with loss of ER expression (Chou et al. 2010; Mehra et al. 2005; Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003). Notably, GATA3 is mutated in more than 10% of all breast cancers. These mutations are markedly enriched in luminal cancers, which are typically ER+, compared to basal cancers, which are typically ER- (~15% versus ~2% of cases) (Cancer Genome Atlas Network 2012). These mutations cluster in the second zinc finger of GATA3, but are diverse and result in varying changes to the overall protein, including both truncations and elongations (Cancer Genome Atlas Network 2012; Emmanuel et al. 2018; Usary et al. 2004). Despite studies on the effects of certain mutants in breast cancer cell lines (Adomas et al. 2014; Gustin et al. 2017; Takaku et al. 2018), the association between mutant GATA3 and clinical outcome is unclear (Du et al. 2015; Liu et al. 2016), and it is possible that the various GATA3 mutations may each have different functional consequences. It has been suggested that GATA proteins may also possess pioneer factor-like activities (Cirillo et al. 2002; Takaku et al. 2016). However, although loss of GATA3 in MCF7 cells does affect ER binding, this appears to be a context-dependent modulation, with equivalent numbers of stronger and weaker ER binding sites observed (Theodorou et al. 2013). Therefore, although GATA3 demonstrates a clear and undeniable connection to ER+ disease, this connection may be complex, and further investigations are required to determine its precise contribution to ER signalling.

1.3.1.3 Tethering proteins

Additional transcription factors often found enriched at ER sites include activator protein 1 (AP-1, a heterodimer consisting of proteins belonging to the c-Jun and c-Fos families) and specificity protein 1 (Sp1). Experiments have shown that these factors likely mediate tethering of ER to the DNA, through the finding that ER deficient in its DNA-binding capacity can still efficiently modulate oestrogen-regulated transcription at ER/AP-1 and ER/Sp1 co-occupied sites (Porter et al. 1997; Jakacka et al. 2001). More recent studies investigating the genome-wide binding profile of an ER DBD mutant have shown that the transcription factor runt-related transcription factor 1 (RUNX1) may also have a role in tethering ER to the DNA at certain regions (Stender et al. 2010).

1.3.2 Co-regulators mediating ER target gene activation and repression

1.3.2.1 Histone modifiers

Histone modifications are highly implicated in the co-ordinated regulation of transcription. Through altering histone-DNA interactions, they primarily serve to regulate chromatin accessibility, modulating the assembly of transcriptional complexes to regulate gene expression in a context-dependent manner. In addition, histone marks may also act as binding sites for transcriptional co-regulators (Zhou et al. 2011; Shlyueva et al. 2014). Different types of histone modifications are associated with gene activation, whilst others correlate with gene repression, and the dynamic regulation of these marks achieved by the ER complex as a whole is diverse and wide-ranging. Focusing on proteins immediately and directly recruited to the activated ER complex, many of these coordinate changes in histone modifications. These include members of the p160 family, such as nuclear receptor co-activators NCOA1 (SRC1), NCOA2 (SRC2 or GRIP1) and NCOA3 (SRC3 or AIB1) (Anzick et al. 1997; Hong et al. 1997; Onate et al. 1995). Whilst these factors do not have intrinsic enzymatic capacity, they act as adapter proteins for the recruitment of histone acetyltransferases (HATs), which modify histone N-terminal tails through the addition of acetyl groups (Rollins et al. 2015). The p160 proteins interact with ER via their central nuclear receptor interaction domain, where LXXLL motifs facilitate interactions with the AF-2 domain of ER (Heery et al. 1997). This provides a platform for the recruitment of HATs such as CREB-binding protein (CBP), p300 and pCAF (p300/CBP-associated factor). These enzymes acetylate histones, which weakens DNA-histone interactions, decondensing chromatin and facilitating the recruitment of further transcriptional co-regulators. As evidence of the importance of this process, histone acetylation, in particular at lysine 27 of histone H3 (H3K27ac) is a key mark associated with active enhancers (Zhou at al. 2011), and increases at many ER-regulated enhancers in response to oestrogen treatment (Hah et al. 2013; Lupien et al. 2009). HATs CBP and p300 can also mediate interactions with active RNA polymerase II, providing a connection to the general transcription machinery (Neish et al. 1998; Rollins et al. 2015).

Additional histone modifiers that are recruited to the activated ER complex and positively regulate transcription include histone methyltransferases (HMTs) such as protein arginine methyltransferases 1 (PRMT1) and 4 (PRMT4, also known as CARM1). Whilst HATs CBP

and p300 are recruited to the ER complex through interactions with the activation domain 1 (AD1) of p160 family proteins, PRMT1 and CARM1 can be recruited through a separate activation domain, AD2 (Koh et al. 2001). Through methylating arginine residues of histones H3 and H4, CARM1 and PRMT4 act synergistically with CBP and p300 to decondense chromatin and enhance ER-mediated transcription (Chen et al. 2000; Ma et al. 2001; Wagner et al. 2006).

Whilst ER is most often discussed in the context of gene activation, about half of ER-regulated genes are repressed in response to oestrogen (Carroll et al. 2006). The precise mechanisms that determine whether ER represses or activates transcription remain unclear. It has been suggested that ER binding correlates less strongly with the TSSs of oestrogen-downregulated genes than with the TSSs of oestrogen-upregulated genes, implying that ER-mediated repression may more frequently involve distal or secondary regulation compared to ER-mediated induction (Carroll et al. 2006). Mechanistically, it is likely that ER achieves gene repression through a variety of means. This may include indirect methods such as displacement of co-activators, or rate-limiting competition for co-activators, known as squelching. However, ER is also known to interact with a number of transcriptional co-repressors, implying a role for ER in mediating direct oestrogen-stimulated repression of some of its target genes (Zubairy & Oesterreich, 2005). Similarly to ER-mediated induction, this ER-mediated repression is also thought to involve changes in histone modifications, for example through adapter-mediated recruitment of histone deacetylases (HDACs). These enzymes counter the effects of HATs by removing activating histone acetylation marks, leading to a more condensed chromatin state. Receptor-interacting protein 140 (RIP140), repressor of estrogen receptor activity (REA) and the nuclear receptor co-repressor (NCOR) family proteins are some of the cofactors that can direct ER-mediated gene repression through their interactions with HDACs (Castet et al. 2004; Delage-Mourroux et al. 2000; Lazar 2003; Stossi et al. 2006; Varlakhanova et al. 2010). As many of these co-repressors, like the p160 co-activators, also interact with ER via its AF-2 domain, co-activator and co-repressor recruitment are often mutually exclusive (Watson et al. 2012). This implies that competition between functionally opposing co-regulators can contribute to the distinct activating or repressing actions of ER, and indeed it has been shown that REA and NCOA1 compete for binding to the ER AF-2 domain (Delage-Mourroux et al. 2000). Nevertheless, certain HDACs may also be capable of interacting with ER directly (Leong et al. 2005).

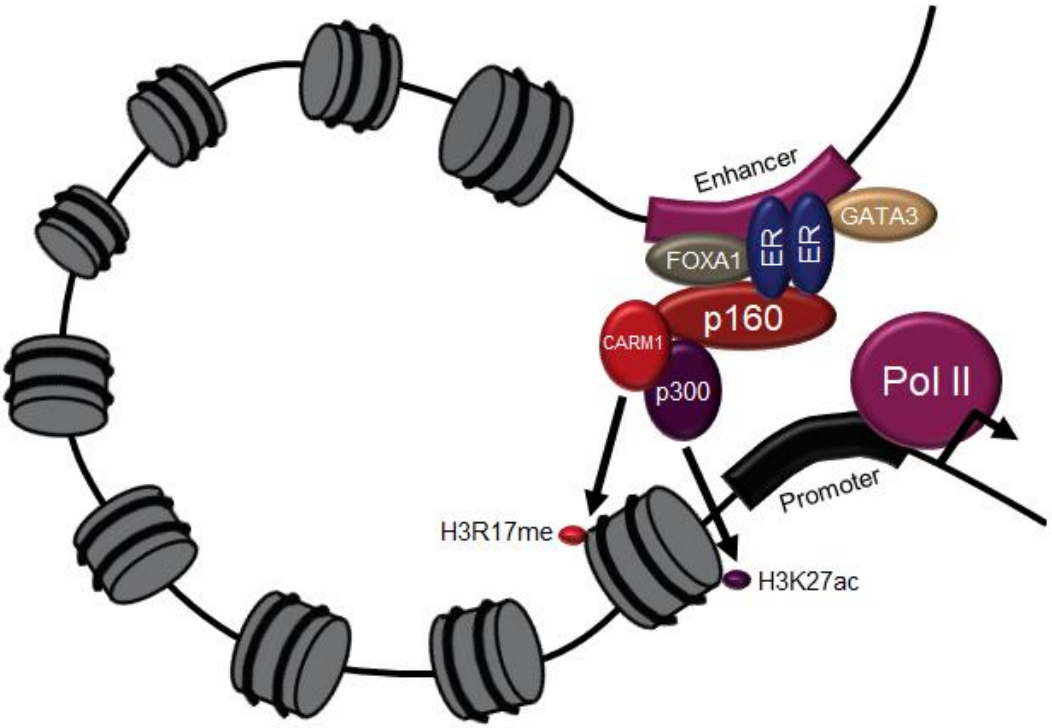


Figure 1.2. The ER complex recruits co-activators regulating histone modifications. ER binds at enhancers in conjunction with other key transcription factors, such as the pioneer factor FOXA1, and ER-associated protein GATA3. The recruitment of p160 adapters to the activated ER provides a platform for interactions with histone acetyltransferases such as p300, and histone methyltransferases such as CARM1. The histone modifications mediated by these enzymes decondense chromatin and facilitate activation of transcription by RNA polymerase II (Pol II). H3K27ac = acetylation of histone H3 lysine 27; H3R17me = methylation of histone H3 arginine 17.

1.3.2.2 ATP-dependent chromatin remodellers

In addition to the changes in chromatin structure induced by ER-directed histone modifications, ATP-dependent chromatin remodelling complexes also have a role in oestrogen-activated gene expression. These complexes alter the spatial organisation of nucleosomes. This is achieved either through repositioning nucleosomes along DNA, restructuring them through exchanging their histone components for different variants, or mediating their expulsion or disassembly (Wang et al. 2007). All of these processes may render chromatin more or less accessible to transcription factors. Brahma-Related Gene 1 (BRG1) and BRG1-associated factor 57 (BAF57), both members of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodelling complex, have been shown to

interact with ER via its AF-2 domain, and appear important for oestrogen-mediated gene expression (Belandia et al. 2002; DiRenzo et al. 2000; García-Pedrero et al. 2006). More specifically, DiRenzo et al. (2000) showed that BRG1-mediated co-activation of ER activity is dependent on its ATPase function, indicating that active chromatin remodelling is a key component of ER transcriptional regulation.

1.3.2.3 DNA methylation

The chemical composition of DNA itself can also undgero epigenetic changes, and it appears that DNA modifications may also play a role in ER-regulated transcription. Methylation of the 5' carbon of cytosine by DNA methyltransferases (DNMTs) results in 5-methylcytosine (5mC), an epigenetic mark thought to play a role in transcriptional regulation (Moore et al. 2013). Several studies have demonstrated a link between ER signalling, methylation, and transcriptional regulation at specific loci. For example, oestrogen treatment induces methylation of the *CYP1A1* promoter and silencing of CYP1A1 expression in MCF7 cells (Marques et al. 2013). In a broader study, Ariazi et al. (2017) proposed ER may repress a subset of its target genes, including tumour suppressors and genes relating to basal-like phenotypes, through directing methylation of regulatory sites. In contrast, studies by Leu et al. (2004) and Stone et al. (2012) both showed that dysregulation of ER signalling (via silencing or antihormone treatment, respectively) resulted in downregulation of the ER target gene *PGR*. This could not be reversed by reactivation of ER signalling alone, but also required demethylation of the *PGR* promoter via DNMT inhibition, implying that loss of ER activity may promote methylation-induced silencing. More recently, Wang et al. (2018) similarly demonstrated that ER-directed DNA demethylation may have a role in oestrogen-mediated gene expression. Finally, Kangaspeska et al. (2008) and Métivier et al. (2008) have shown in MDA-MB-231 cells stably expressing ER that the promoter of *TFF1* (*pS2*), an oestrogen-regulated gene, undergoes cyclical changes in methylation in response to oestrogen. Blocking this process inhibited TFF1 expression, implying that whilst stable changes in DNA methylation might contribute to ER transcriptional control, the dynamics of these changes may also be important in defining their effects.

1.4 Targeted therapies for ER positive breast cancer

Given that ER is the driving transcription factor in the majority of breast cancers, numerous treatment strategies centre around targeting either ER itself, or oestrogen as its activating ligand. Three main classes of these drugs, discussed here, are selective oestrogen receptor modulators (SERMs) such as tamoxifen, selective oestrogen receptor degraders (SERDs) such as fulvestrant, and aromatase inhibitors (AIs). Whilst SERDs act as pure antioestrogens and have only antagonist activity in relation to ER, SERMs are distinguished from both full ER agonists and antagonists in that they may have antioestrogenic activity in some tissues whilst providing oestrogenic activity in others. In contrast, AIs act upstream of ER by blocking oestrogen synthesis.

1.4.1 Tamoxifen

Originally and unsuccessfully designed as a contraceptive pill, a role for tamoxifen in treating breast cancer was first identified in 1971 (Cole et al. 1971). Tamoxifen has since been in use for more than 40 years for the effective treatment of ER+ disease (Jordan 2003). Illustrating the success of this drug, recent analyses have shown that five years' tamoxifen treatment halves the 10-year recurrence rate for ER+ breast cancer, corresponding to an approximately one-third decrease in mortality (Early Breast Cancer Trialists' Collaborative Group (EBCTCG) et al. 2011). In breast cancer, the actions of tamoxifen are two-fold. Firstly, tamoxifen competes with endogenous oestrogens for binding to the ER LBD. Secondly, tamoxifen induces conformational changes in ER that impair its association with co-activating p160 proteins, and consequently with transcriptional co-activators such as p300 and CBP. This promotes the interaction of ER with co-repressors such as SMRT and NCOR and blocks the expression of genes driving tumour growth (Brzozowski et al. 1997; Shang et al. 2000; Shang & Brown 2002; Shiau et al. 1998). Whist tamoxifen antagonises ER activity in breast, it can activate ER in other tissues, for example in bone, hence its definition as a SERM (Haskell 2003). These tissue-specific effects are thought to be due to varying local levels of endogenous oestrogens, and the differing expression of ER co-activators and co-repressors in different cell types (Dutertre & Smith 2000).

1.4.2 Aromatase inhibitors

Aromatase inhibitors (AIs) affect ER activity through blocking oestrogen production. Oestrogen is synthesised through conversion from androgen by aromatase, a member of the cytochrome P450 (CYP)19 family of enzymes (Cole & Robinson 1990). AIs in use in the clinic include letrozole, anastrazole, and exemestane, all of which interact competitively with the active site of aromatase and prevent androgen binding (Miller 2003). AIs are most appropriate for use in postmenopausal women where oestrogen production switches from the ovaries to peripheral tissues including the breast. This is due to the fact that in premenopausal women, not only are circulating oestrogen levels much higher, but increased hormone synthesis by the ovaries can rapidly counteract any oestrogen-depleting effects of AIs (Simpson 2003). In contrast, local production of oestrogen can be effectively lowered by aromatase inhibition in the postmenopausal setting, facilitating reduced activation of ER.

1.4.3 Fulvestrant

In contrast to tamoxifen, fulvestrant does not demonstrate any oestrogenic effects, and uniformly abrogates ER activity. Its mechanism of action involves high-affinity binding to ER leading to conformational changes that inhibit ER dimersation and nuclear translocation. In addition to blocking ER transcriptional activity, fulvestrant also triggers accelerated degradation of ER, resulting in a rapid depletion of total ER levels (Dauvois et al. 1993; Fawell et al. 1990; Wakeling et al. 1991). Due to its orthogonal mode of action, fulvestrant can be used as a second-line treatment in breast cancers exhibiting intrinsic or acquired tamoxifen resistance. However, unlike tamoxifen and AIs, fulvestrant cannot be given orally and is administered by intramuscular injection in the clinic. Although this can aid treatment compliance, this renders fulvestrant treatment more expensive than tamoxifen or AIs. Given that the performance of AIs and fulvestrant are comparable as second-line therapies, AIs are therefore typically favoured for treatment of postmenopausal women (Robertson 2007).

1.5 TET proteins

Ten-eleven translocation (TET) proteins were first described with the identification of *TET1* as a fusion partner of the mixed lineage leukaemia (*MLL*) gene in acute myeloid leukaemia (Lorsbach et al. 2003). Subsequent investigations revealed two additional TET family members, TET2 and TET3. Through their homology to the trypanosome enzymes J-binding protein 1 and 2 (JBP1 and JBP2), TET proteins were determined as Fe(II)/α-ketoglutarate-dependent dioxygenases capable of modifying methylated cytosine (5mC) (Tahiliani et al. 2009). Specifically, TETs oxidise 5mC, and in an iterative process, produce the further DNA modifications 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009), and the implication of 5mC in genome stability and transcriptional control has driven interest in TET enzymes as dynamic regulators of this epigenetic mark. Emerging evidence that the 5mC oxidation products 5hmC, 5fC and 5caC may also have transcriptional roles in their own right (discussed in section 1.7 below) has further fuelled research into TET proteins, primarily in development, but also more recently, in cancer.

1.5.1 Structure of TET proteins

TET proteins are large (~180 to 230 kDa) multidomain enzymes. Their shared C-terminal catalytic domain consists of a double stranded beta helix (DSBH) domain, binding sites for Fe(II) and α-ketoglutarate cofactors, and a cysteine-rich (Cys) domain (Figure 1.3). Whilst the conserved catalytic domain mediates direct interactions with cytosine residues, this does not define the wider DNA sequence preference of these enzymes (Hu et al. 2013). TET1 and TET3 possess a CXXC domain which facilitates their targeting to CG-rich regions (Xu et al. 2018). However, TET2 lacks this domain, hence the mechanisms by which it is targeted to specific genomic regions remain unclear (Rasmussen & Helin 2016). It is possible that the less-conserved regions of TET2 provide interfaces for interactions with different proteins that may direct its sequence-specific genomic targeting (Hu et al. 2013). This basic structural difference provides one feature that distinguishes TET2 from its other family members. The remainder of this section will discuss the overall functions of TET proteins, with a focus on TET2.

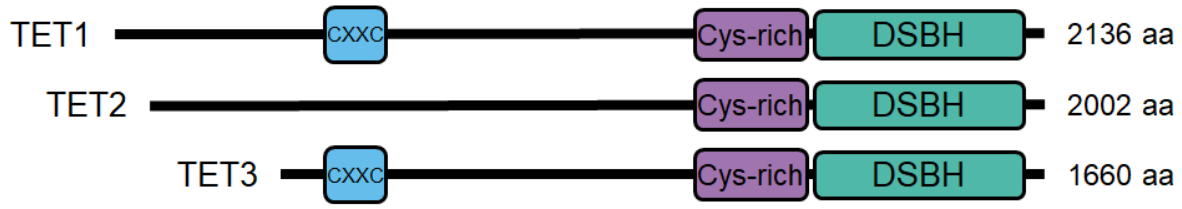


Figure 1.3. Structure of TET proteins.
The core catalytic domain of all three TET enzymes consists of a cysteine-rich (Cys-rich) domain, a double stranded beta helix (DSBH) domain, and binding sites for Fe(II) and α-ketoglutarate cofactors (within the DSBH, not shown). Additionally, TET1 and TET3 possess an N-terminal CXXC domain facilitating their sequence-specific DNA binding.

1.5.2 TET2 in development and disease

Whilst homozygous deletion of *TET3* is embryonic lethal in mice (Gu et al. 2011), deletion of *TET1* or *TET2* causes no striking developmental phenotypes, but rather skews the differentiation of certain cell lineages. With respect to loss of *TET2*, this altered cellular differentiation occurs specifically in the haematopoietic compartment (Dawlaty et al. 2011; Moran-Crusio et al. 2011; Quivoron et al. 2011). In contrast, combined deletion of *TET1* and *TET2* in mice results in a more severe phenotype, and most die perinatally (Dawlaty et al. 2013). Therefore, whilst TET3 appears to have a unique role in embryonic development, TET1 and TET2 may be capable of compensating for one another in this setting.

The important role of TET2 in regulating haematopoetic development becomes clearer through studies of its involvement in cancer, where it is the most frequently mutated gene in a range of haematological malignancies, including chronic myelomonocytic leukaemia (CMML, *TET2* mutated in 20-58% of cases), acute myeloid leukaemia (AML, 12-32% of cases), and T-cell lymphomas (20-83% of cases). Across these malignancies, mutations of *TET2* appear varied, encompassing missense mutations that result in various changes to the amino acid sequence, nonsense mutations causing truncations, frameshift deletions/insertions, and also splice site mutations. These events do not localise to a specific exon or residue, and there appears no predominant hotspot for these alterations. In addition, many patients exhibit more than one *TET2* mutation. Overall, this evidence

implies TET2 inactivation is the unifying outcome of these mutational occurrences (Kosmider et al. 2009; Patnaik et al. 2016). These haematological TET mutations appear to be specific to *TET2*, as *TET1* and *TET3* are rarely found mutated in these cancers (Scourzic et al. 2015).

In contrast to their well-documented role in haematological malignancies, less attention has been drawn to the oncogenic potential of *TET2* mutations in solid cancers. *TET2* mutations can be identified in many solid tumour types according to data available through the online tool CBioPortal (www.cbioportal.org) (Cerami et al. 2012; Gao et al. 2013). However, with the exception of a few studies showing moderate levels of *TET2* perturbation in non-small cell lung cancer (Gardner et al. 2017) and skin cancer (Pickering et al. 2014), mutational frequencies of *TET2* in solid cancers tend to be low, at between 0.1 – 10%, and *TET2* mutations are rare in breast cancer (Scourzic et al. 2015; Stephens et al. 2012). Of the infrequent mutations observed in solid tumours, in a similar manner to the mutations observed in haematological malignancies, these alterations do not localise to one specific region of the protein, and the associated changes to the protein sequence vary. Overall, missense mutations appear to be the most common type of alteration to *TET2* in solid tumours, with truncating mutations second most common. A small number of splice site mutations can also be observed in the solid tumour datasets available on CBioPortal, and TET2 exhibits fusions with several different loci in a small minority of studies. Nevertheless, despite its low mutational frequency, reduced TET2 expression is observed in several cancer types, including endometrial, liver and breast cancer (Ciesielski et al. 2017; Yang et al. 2013). Linking TET2 levels to clinical outcome, Takayama et al. (2015) showed that loss of TET2 is associated with metastasis in prostate cancer, and in a mixed cohort of 162 breast cancer patients, Yang et al. (2015) demonstrated reduced expression of TET1, TET2, and TET3 mRNAs as associated with poor prognosis. Of note, Huang et al. (2016) reported loss of TET2 nuclear localisation (although not overall expression) as associated with a more aggressive phenotype in colorectal cancer, implying that exclusion of TET2 from the nucleus may play a role in this malignancy.

1.5.3 TET2 functions

1.5.3.1 Catalytic functions of TET2

In terms of its catalytic functions, TET2, like TET1 and TET3, iteratively oxidises 5mC to create the further DNA modifications 5hmC, 5fC and 5caC (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009). 5mC is initially formed through methylation of cytosines by DNMTs. This typically occurs in the context of CpG dinucleotides, self-complementary DNA sequences where a cytosine is followed by a guanine on the same strand. During cell division, 5mC can be maintained through the recognition of newly replicated, hemi-methylated DNA by ubiquitin-like PHD and RING finger domain-containing protein 1 (UHRF1), which promotes recruitment of the maintenance methyltransferase DNMT1 to methylate the newly-synthesised strand in a symmetrical manner (Bostick et al. 2007; Hermann et al. 2004). Where UHRF1/DNMT1 do not act to reinstate 5mC in this manner, for example through reduced activity of DNMT1, this can result in passive, replication-dependent loss of 5mC.

The effect that TET-mediated oxidation of 5mC has on the heritability of this mark has been studied. Although the DNMT1 interaction partners UHRF1 and UHRF2 can also interact with 5hmC (Frauer et al. 2011; Hashimoto et al. 2012; Iurlaro et al. 2013; Spruijt et al. 2013; Zhou et al. 2014), the activity of DNMT1 at hemi-hydroxymethylated DNA appears to be up to 60-fold less than its activity at hemi-methylated DNA (Hashimoto et al. 2012; Ji et al. 2014; Valinluck & Sowers 2007). This has led to the suggestion that TET-mediated conversion of 5mC to 5hmC may trigger passive loss of DNA modifications, and indeed, this process appears to be important both in germ cells and in early embryogenesis (Hackett et al. 2013; Inoue & Zhang 2011). However, alternative methyltransferases DNMT3A2 and DNMT3B2 may also be involved in mitotic inheritance of 5mC, and these enzymes are also capable of methylating hemi-hydroxymethylated sites (Hashimoto et al. 2012; Ji et al. 2014). This suggests that the presence of 5hmC may not always result in replication-dependent loss of 5mC. Further investigations will be necessary to determine the effect of TET-mediated 5hmC generation on the maintenance of DNA modifications, however it is possible that this effect may be both cell context- and locus-specific.

No inheritance mechanism has been described for the final products of TET metabolism, 5fC and 5caC, thus it appears likely that these marks can also act as intermediates in TET-mediated, replication-dependent DNA demethylation. However, in contrast to 5hmC, these later oxidised products can also be actively removed from DNA. Specifically, both 5fC and 5caC are substrates for thymine DNA glycosylase (TDG), an enzyme typically mediating DNA mismatch repair (He et al. 2011; Maiti & Drohat 2011; Spruijt et al. 2013). TDG-mediated excision of its DNA substrates results in the generation of an abasic site which is subsequently replaced with an unmodified cytosine through base excision repair (BER) pathways (Lindahl & Wood 1999). In support of this mechanism operating to remove 5fC and 5caC, disruption of TDG in embryonic stem cells (ESCs) causes accumulation of both of these marks (He et al. 2011; Shen et al. 2013; Song et al. 2013), and overexpression of TDG in HEK293T cells results in their depletion (Nabel et al. 2012). Notably, TDG displays no *in vitro* activity towards unmodified cytosine, 5mC or 5hmC, and overexpression of TDG appears to have no effect on 5mC and 5hmC levels, indicating that these active demethylation pathways are specific to 5fC and 5caC (He et al. 2011; Maiti & Drohat 2011; Nabel et al. 2012) Overall, this indicates a key role for TET enzymes in DNA demethylation pathways. A summary of these pathways is shown in Figure 1.4.

Of the TET-regulated DNA modifications, 5mC and 5hmC are the most abundant. 5mC is typically present as approximately 4-5% of total cytosines, whilst levels of 5hmC range between 0.03-0.7% due to wide tissue-specific variation (Globisch et al., 2010; Lister et al. 2009; Szwagierczak et al. 2010). Contrastingly, 5fC and 5caC are present at levels at least 1000 times lower than 5hmC (Globisch et al. 2010; Ito et al. 2011), and are more difficult to reliably detect using current methods. The scarcity of 5fC and 5caC in comparison to 5hmC further reinforces that these marks may be actively maintained at low levels in the genome through TET/TDG-mediated removal. However, the relative abundance of 5hmC may also be due to more efficient production of this mark compared to 5fC and 5caC. Specifically, TET-mediated conversion of 5hmC to 5fC appears to be ~6-fold slower, on average, than the initial conversion of 5mC to 5hmC, whilst the rate of oxidation of 5fC to 5caC is ~10-fold reduced compared to the initial 5mC to 5hmC step (Hu et al. 2015; Ito et al. 2011; Rasmussen et al. 2016). This suggests that TET enzymes stall in their activity after the production of 5hmC. The resultant relative stability of 5hmC compared to 5fC and 5caC could indicate a unique cellular role for this mark. Additionally, it has been proposed that the main function of all three TET proteins may be to prevent

aberrant methylation of non-methylated regions (Jeong et al. 2014; Rasmussen et al. 2015). Therefore, the finding 5mC appears to be the mark most efficiently converted by TET proteins may also support the notion of 5mC suppression as the most important biological role of these enzymes.

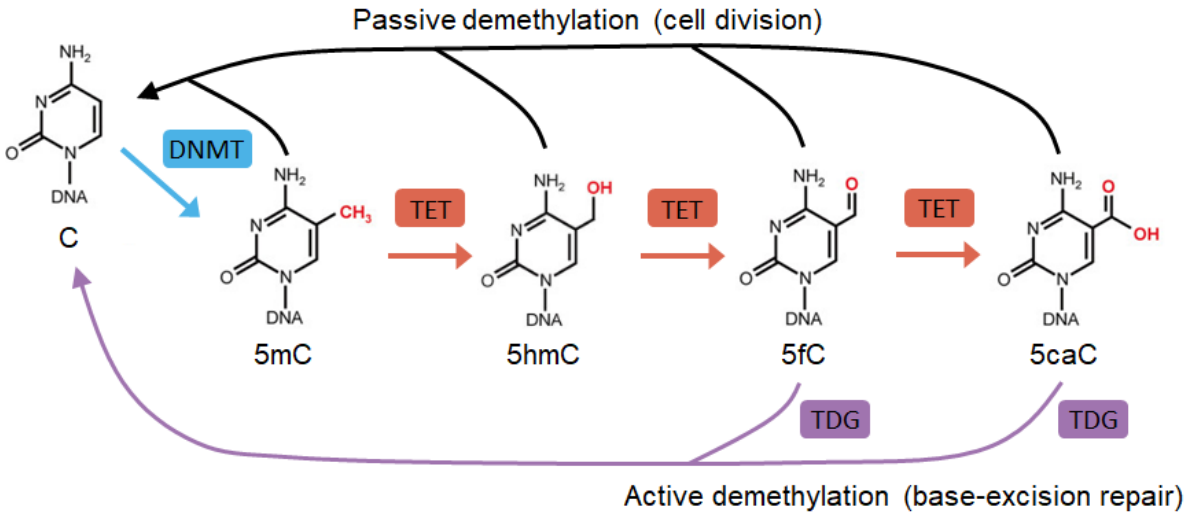


Figure 1.4. The role of TET enzymes in active and passive DNA demethylation.
Cytosine residues are methylated at the 5' carbon by DNA methyltransferases (DNMTs), resulting in 5-methylcytosine (5mC). This newly-added methyl group can be further modified by the addition of a hydroxyl group by TET enzymes to generate 5-hydroxymethylcytosine (5hmC). 5hmC can be further oxidised by TET enzymes to produce, sequentially, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Whilst all these modifications may be lost passively through DNA replication as a result of cell division, 5fC and 5caC can also be removed through base excision repair pathways initiated by thymine DNA glycosylase (TDG).

Studies in numerous different systems have demonstrated the key role of TET enzymes in regulating DNA modifications. Notably, despite the fact that the basic catalytic functions of all three TET proteins are broadly similar, several investigations have demonstrated that these enzymes may preferentially regulate modifications at different sites, with a role for TET2 that appears distinct from the other TET family members. For example, Putiri et al. (2014) showed that siRNA-mediated depletion of TETs 1, 2, and 3 in an embryonic carcinoma cell line indicated a clear role for all three family members in regulation of 5hmC at enhancers. However, of the three TET proteins, TET2 appeared to mediate the most robust regulation of this mark at these sites. This finding was corroborated by Hon et al.

(2014) in mouse embryonic stem cells (mESCs), who showed that loss of both TET1 and TET2 resulted in accumulation of 5mC at a range of sites. Whilst the vast majority of these hypermethylated sites encompassed enhancers and distal regulatory elements (constituting 70% of hypermethylated sites in response to TET1 loss, and 86% of sites in response to TET2 loss), the absolute number of sites affected by TET2 was approximately 8.7 times the number of sites hypermethylated in response to TET1 loss (60,095 versus 8,455 sites). In the same study, measurements of 5hmC similarly demonstrated that whilst loss of both TET1 and TET2 resulted in depletion of this mark at promoters, gene bodies, CTCF-bound insulators, and enhancers, the effects of TET2 loss on 5hmC were more pronounced that those of TET1 loss. Therefore, whilst knockout studies imply that TET1 and TET2 may be capable of overlapping effects in the developmental setting, functional investigations in mESCs and cancer cell lines show they have distinct roles at the level of DNA modifications. Furthermore, it appears TET2 may possess a key role in regulating DNA modifications at enhancers.

1.5.3.2 Additional functions of TET2

In addition to the catalytic role of TET2, the transcriptional implications of which are described in section 1.7 below, several further roles have been described for TET proteins that may not rely solely on their dioxygenase capabilities. For example, whilst perturbation of TET catalytic activity has an impact on the levels and distribution of 5mC and its oxidised forms, it has been suggested that TET proteins may also prevent aberrant 5mC deposition at selected genomic regions through physical occupancy of these sites (Wu & Zhang 2017). Citing a study by Lu et al. (2014), which investigated methylation patterns in mESCs in response to knockout of all three TET proteins, Wu & Zhang (2017) proposed that the accumulation of 5mC at sites where oxidised forms of 5mC are normally present at very low levels implies TET proteins could partly prevent aberrant methylation at these sites through blocking access to DNMTs.

TET proteins may also have a role in stabilising or modifying the activity of transcriptional complexes, potentially acting as an adapter protein, through associating with histone modifying enzymes. For example, all three TET proteins can directly interact with the enzyme O-linked β-N-acetylglucosamine transferase (OGT) (Chen et al. 2013; Deplus et al. 2013; Vella et al. 2013). OGT can modify histones through *O*-GlcNAcylation, and has

also been linked to processes mediating histone ubiquitination (Deplus et al. 2013; Hardivillé & Hart 2014; Lercher et al. 2015). In addition, TET2/OGT interactions have been shown to be important for stabilisation of the histone methyltransferase complex SET1/COMPASS, with a study by Deplus et al. (2013) demonstrating that the interaction between TET2 and OGT is important for H3K4me3 deposition and transcriptional activation in HEK293T cells. These examples provide a potential connection between TET enzymes and wider epigenetic functions.

Furthermore, several studies have demonstrated that TET catalytic activity may be negligible for some of its functions. Specifically to TET2, Zhang et al. (2015) showed that TET2-mediated HDAC2 recruitment can repress interleukin 6 (IL-6) expression in mouse myeloid cells, and through depletion of wild-type TET2 and overexpression of a TET2 catalytic mutant, the authors suggested this may be independent of its enzymatic function. Similarly, Montagner et al. (2016) showed that in mast cells, hyperproliferation resulting from TET2 loss could be rescued by re-expression of either wild-type or catalytically inactive TET2.

In terms of the other TET family members, TET1 interacts with the SIN3A histone deacetylase complex (Williams et al. 2011), and the histone modifier lysine (K) acetyltransferase 8 (KAT8) (Zhong et al. 2017). Additionally, TET3-mediated repression of the imprinted *SNRPN* gene, required for preventing premature differentiation of mouse NSCs, is maintained even in the presence of a TET3 catalytic mutant (Montalbán-Loro et al. 2019). These recent works imply that further and potentially diverse roles of TET proteins remain to be uncovered.

1.6 DNA modifications in transcriptional control

Due to the role of TET enzymes in regulating the levels of 5mC and its oxidised forms, the distribution and function of these epigenetic marks is thought to be key to the biological importance of TET proteins. The next two sections will provide a brief overview of the suspected role of 5mC and its TET-oxidised intermediates in transcription, touching upon the potential dysregulation of these marks in cancer.

1.6.1 Distribution and proposed transcriptional roles of DNA modifications

The role of DNA methylation in transcriptional regulation has been investigated in a number of systems, with the broad conclusion being that this mark is associated with transcriptional repression (Bird & Wolffe 1999; Bird 2002; Klose & Bird 2006; Moore et al. 2013). DNA methylation occurs primarily at cytosines in the context of CpG dinucleotides. Non-CpG modification is also possible, although this is a rarer event (Lister et al. 2009). The frequency of CpGs in the genome is overall lower than would be expected by chance, partly due to the tendency of 5mC to mutate into thymine. However, the majority of CpG dinucleotides (70-80%) throughout the mammalian genome are methylated (Bird 2002). DNA methylation is associated with various processes, including X chromosome inactivation and genomic imprinting, and a large proportion of DNA methylation is thought to be involved in repression of endogenous retrotransposons and viral elements (Beard et al. 1995; Jaenisch et al. 1985; Mohandas et al. 1981; Schulz et al. 2006). Of key relevance to the role in 5mC in transcription, most (60-70%) promoters contain CpG islands (CGIs). These are stretches of DNA that contain a higher frequency of CpG dinucleotides than the surrounding genome, but are largely protected from DNA methylation (Illingworth et al. 2010). Whilst most promoters are therefore depleted of 5mC irrespective of the transcriptional activity of the gene they control, where methylation of CGI-containing promoters does occur, this is negatively correlated with gene expression (Bird & Wolffe, 1999).

Two main mechanisms have been proposed for 5mC-directed transcriptional repression. Firstly, the presence of 5mC is thought to have a destabilising influence on the DNA binding of most transcription factors and DNA-binding proteins, hindering transcriptional activation (Yin et al. 2017). Secondly, of the proteins shown to interact directly with 5mC, the earliest of these identified were transcriptional repressors, which recruit histone deacetylase complexes and catalyse chromatin condensation (Bird & Wolffe 1999; Jones et al. 1998; Nan et al. 1998). Examples of these methyl-CpG-binding domain (MBD)-containing proteins include MeCP1, MeCP2, and MBD1, 2, and 4 (Klose & Bird 2006). In addition, the transcriptional repressor Kaiso (ZBTB33) has been shown to bind methylated CpG dinucleotides through its zinc-finger domains (Prokhortchouk et al. 2001; Yoon et al.

2003), and it is possible that other members of this large protein family may possess similar functions (Filion et al. 2006).

Nevertheless, despite this mechanistic evidence for the repressive activity of 5mC, recent studies have identified a range of additional proteins capable of interacting with 5mC, not all of which are associated with transcriptional repression, and which may have more diverse regulatory functions (Iurlaro et al. 2013; Spruijt et al. 2013; Spruijt & Vermeulen 2014). In addition, the fact that a strong transcriptional activator can trigger transcription from a previously silenced, methylated promoter (Thompson et al. 1986), and that genes possessing methylated CGIs are expressed in some developmental contexts (Fouse et al. 2008), implies that a balance between 5mC and the profile of transcriptional regulators that interact with this mark may “fine-tune” gene expression at different loci. It therefore appears that further work is required to precisely define the potentially complex relationship between 5mC and gene expression.

In terms of levels of 5mC at other regulatory elements, overall, enhancers also appear to be depleted of this mark, although less so than promoters (Luo et al. 2018; Stadler et al. 2011), and at these sites 5mC loss is similarly correlated with positive transcriptional activity (Hon et al. 2013; Lister et al. 2009; Stadler et al. 2011; Ziller et al. 2013). Importantly, Ziller et al. (2013) examined genome-wide DNA methylation across 30 human cell and tissue types through investigation of 42 separate whole genome bisulfite sequencing (WGBS) datasets. This analysis showed that dynamic regulation of methylation occurred at only 21.8% of CpG sites within a normal developmental context. The majority of these dynamic CpGs co-localised with gene regulatory elements, particularly enhancers and transcription factor binding sites linked to cell type-specific genes. This indicates that regulation of 5mC at enhancers may be a key process in the definition of cell phenotypes, with interplay between DNMT-mediated methylation and TET-mediated demethylation potentially central to this. As the studies cited in the previous section highlight, TET depletion in several systems corroborates this notion, with the finding that regulatory elements appear to be the main targets of TET-mediated regulation of DNA modifications (Hon et al. 2014; Lu et al. 2014; Stadler et al. 2011).

In contrast to 5mC, which is notable in its distribution by the regions from which it is excluded, 5hmC is specifically enriched at various genomic regions, including promoters, gene bodies, and enhancers (Hahn et al. 2013; Johnson et al. 2016; Pastor et al. 2011; Stroud et al. 2011; Wu et al. 2011). Whilst studies show 5hmC at enhancers is typically associated with transcriptional activation, its role at promoters remains unclear, with several studies proposing 5hmC as an activating mark at these sites (Hahn et al. 2013; Madzo et al. 2014; Taylor et al. 2016), and others suggesting it may have a repressive role (Neri et al. 2013; Wu et al. 2011). However, as 5mC is required as a substrate for conversion to 5hmC, these conflicting findings may be partly due to different basal levels of 5mC at these promoters that confound measurement of the effects of 5hmC (Pastor et al. 2013). Interestingly, screens conducted by Iurlaro et al. (2013) and Spruijt et al. (2013) to identify readers of 5mC and its oxidised derivatives demonstrated that 5hmC interacts both with proteins thought to be involved in active DNA demethylation pathways, and those involved in transcriptional regulation. This reinforces the idea of 5hmC as a stable transcriptional regulatory mark, in addition to being a potential DNA demethylation intermediate. Furthermore, in these studies, 5mC and 5hmC were found to attract the binding of distinct proteins, with very few proteins capable of interacting with both marks. This further fuels the notion of distinct transcriptional roles for these two modifications.

Compared to 5mC and 5hmC, the specific functions of the final products of TET metabolism, 5fC and 5caC, have been much less investigated. Nevertheless, the *in vitro* screen performed by Iurlaro et al. (2013) demonstrated that far more specific readers bound uniquely to 5fC than to 5mC or 5hmC, and the functions of these proteins strongly suggest transcriptional regulatory capabilities. Moreover, in a recent study by the same group, genome-wide profiling of 5fC was performed in mouse embryos, through biotin labelling of 5fC followed by pull-down and sequencing (Iurlaro et al. 2016). This showed that 5fC is enriched at active enhancers in mouse embryos in a tissue-specific manner, suggesting a role for this mark in embryonic development. Work in additional systems, and the development of further methods to profile both 5fC and 5caC, may reveal further insights into the potential transcriptional roles of these products of TET activity.

1.6.2 Dysregulation of DNA modifications in cancer

In addition to their role in normal cells, levels of both 5mC and 5hmC may be dysregulated during cancer progression. Many cancers demonstrate a global loss of both 5mC (Hernandez-Blazquez et al. 2000; Kulis & Esteller 2010; Paz et al. 2002; Wilson et al. 2007) and 5hmC (Lian et al. 2012; Kudo et al. 2012; Takayama et al. 2015; Yang et al. 2013), compared to normal tissue. Despite this, increases in 5mC are observed at some sites, and it has long been suggested that hypermethylation of otherwise non-methylated tumour suppressor promoters may contribute to their silencing in cancer (Baylin & Jones, 2016). However, more recent reports have suggested that many of the promoters hypermethylated in cancer cells are silenced in non-malignant tissues to begin with (Sproul et al. 2012), and that hypermethylation of enhancers may instead be a more dominant feature of cancer progression (Bell et al. 2016; Taberlay et al. 2014). Of specific relevance to breast cancer, several studies have clustered breast tumours based on their differential distribution of 5mC and shown that this stratifies tumours based on ER status (Fackler et al. 2011; Hill et al. 2011; Ronneberg et al. 2011).

Changes in 5hmC have been less extensively mapped in normal versus cancer tissues, however local gains in 5mC are often associated with 5hmC loss, with Thomson et al. (2016) specifically showing depletion of 5hmC concomitant with hypermethylation of certain promoters during malignant progression in a mouse model of liver carcinogenesis. Broadly speaking, whether 5hmC loss is a cause or consequence of malignancy remains a topic of debate (Ficz & Gribben 2014). Nevertheless, this loss of 5hmC often correlates with reduced expression of TET proteins (Kudo et al. 2012; Lian et al. 2012; Yang et al. 2013). Overall, it appears that the role of DNA modifications may be diverse and complex, and in particular, examination of the different transcriptional contributions of 5mC versus its oxidised intermediates requires significant work. It appears that TET proteins, and TET2 in particular, may provide a link between these DNA modifications and gene regulatory events at enhancers.

1.6.3 Aims of the thesis

Rationale: ER is a transcription factor that is central to breast cancer biology, and constitutes the main therapeutic target in ER+ disease. Although several targeted

therapies exist for ER+ breast cancer, in many cases patients incur resistance and relapse on these treatments. ER co-operates with numerous other proteins that together form the ER transcriptional complex. The proteins that co-assemble with the activated ER are diverse in function, and their interplay facilitates the precise and dynamic actions of ER in regulating gene expression dependent on the cellular context. The overarching aim of this thesis is to investigate ER signalling in the context of ER+ breast cancer, with the intention of developing a better understanding of how ER functions in the disease setting, primarily in terms of its interaction with cofactors that help it achieve its function.

Aim 1: The transcription factor GATA3 is highly expressed, and frequently mutated, in ER+ breast cancers (Cancer Genome Atlas Network 2012; Kouros-Mehr et al. 2006; Perou et al. 2000; Sorlie et al. 2001). Oestrogen-induced growth of ER+ breast cancer cells appears dependent on GATA3 (Eeckhoute et al. 2007; Kong et al. 2014), and ER and GATA3 share a large proportion (~45%) of ER binding sites in ER+ breast cancer cells (Kong et al. 2011; Theodorou et al. 2013), therefore it appears that there may be functional interplay between these two proteins. The first aim of this thesis constitutes an investigation into the contribution of GATA3 to the ER complex using proteomics-based approaches.

Aim 2: Having established the dioxygenase enzyme TET2 as a key GATA3-regulated component of the ER complex as part of the investigations addressing Aim 1, the second aim of this thesis sets out to further characterise the contribution of TET2 to ER signalling, specifically focusing on the relationship between the ER, GATA3 and TET2 signalling complexes, and the role of TET2 in ER-mediated gene expression and chromatin binding.

Aim 3: The third aim of this thesis expands on the efforts detailed in Aim 2 by specifically focusing on the role of TET2 in regulating DNA modifications in ER+ breast cancer cell lines, and exploring the relationship between these processes and ER-regulated gene expression.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Cell lines and media

ER+ luminal breast cancer cell lines MCF7, T47D and ZR75-1 were obtained from ATCC. MCF7 cells were grown in DMEM (Gibco), T47D cells and ZR75-1 cells in in RPMI 1640 (Gibco). All media was supplemented with 10% heat-inactivated foetal bovine serum (FBS), 50 U/ml penicillin and 50 μg/ml streptomycin (Gibco) and 2 mM L-glutamine (Gibco). Cells were genotyped by short-tandem repeat (STR) profiling using the PowerPlex 16HS Cell Line panel and analysed using Applied Biosystems Gene Mapper ID v3.2.1 software by the external provider Genetica DNA Laboratories (LabCorp Specialty Testing Group) at least every six months and around every major experiment. Cells were routinely mycoplasma tested using MycoProbe Mycoplasma detection kit (R&D).

2.1.2 Patient-derived xenograft (PDX) material

PDX material was kindly provided by Carlos Caldas and colleagues (Bruna et al. 2016). Frozen PDX tissue was propagated in immune-compromised mice. Briefly, tumour pieces (1 mm^3) were implanted into the mammary pad of NSG (NOD scid gamma) mice. All mice were supplemented with oestrogen, using silastic E2 pellets (made in-house) inserted into the dorsal scruff. Tumours were measured twice weekly. Once tumours reached ~1000 mm^3, mice were sacrificed by cervical dislocation under deep, isoflurane-induced anaesthesia. Tumours were resected and either snap frozen in liquid nitrogen, fixed in

10% neutral buffered formalin solution for subsequent paraffin embedding, embedded in Optimal Cutting Temperature compound (OCT), or viably frozen in foetal calf serum (FCS) supplemented with 5% dimethyl sulphoxide (DMSO). STG195 possesses a Y537S mutation in the ESR1 gene; AB555 is ER wild-type.

2.1.3 Antibodies

Antibodies used are listed in Table 2.1.

Protein target	**Antibody**	**Application**
β-actin	Cell Signaling #4970	Western blot
ERα	Novocastra, Leica NCL-L-ER-6F11	Western blot
	Abcam ab3575	ChIP, RIME
	Millipore 06-935	ChIP, RIME
GATA3	Santa Cruz sc268	Western blot, RIME
	Abcam ab106625	RIME
TET2	Abcam ab94580	ChIP, RIME
	Millipore MABE462	ChIP, RIME and Western blot
	Bethyl A304-247A	
	Pierce PA5-3547	
	Santa Cruz sc-398535	
	Cell Signaling 18950	
	Cell Signaling 45010	

Table 2.1. List of antibodies used and their applications.

2.1.4 Small interfering RNAs and drug compounds

Control small interfering RNAs (siRNAs) (D-001810-10), and those used to knock down GATA3 (L-003781-00), ER (L-003401-00), and TET2 (L-013776-03) were obtained from Dharmacon (Horizon Discovery). In order to achieve robust knockdown whilst minimising possible off-target effects of specific siRNA sequences, pools of four different siRNA sequences were used against each target. Sequences are shown in Table 2.2. Cells were transfected with siRNA using Lipofectamine RNAiMax transfection reagent (Invitrogen). For cell treatments, fulvestrant (Selleckchem) was used at a final concentration of 100 nM.

Target	**siRNA sequence**
Non-targeting control	UGGUUUACAUGUCGACUAA UGGUUUACAUGUUGUGUGA UGGUUUACAUGUUUUCUGA UGGUUUACAUGUUUUCCUA
ER	GAUCAAACGCUCUAAGAAG GAAUGUGCCUGGCUAGAGA GAUGAAAGGUGGGAUACGA GCCAGCAGGUGCCCUACUA
GATA3	GUACAGCUCCGGACUCUUC CCCAAGAACAGCUCGUUUA GAAGGCAUCCAGACCAGAA CAUCGACGGUCAAGGCAAC
TET2	ACAAGAAAGUAGAGGGUAU ACACCUAGUUUCAGAGAAU CCUCAGAAUAAUUGUGUGA CAGCAAAGGUACUUGAUAC

Table 2.2. Sequences of siRNAs against ER, GATA3 and TET2.

2.1.5 Primers for qRT-PCR and ChIP-qPCR

Sequences of primers for qRT-PCR are shown in Table 2.3. Ubiquitin C (UBC) was used as a housekeeping control. Sequences of primers for ChIP-qPCR are shown in Table 2.4.

Target	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
TET2	ATTCTCGATTGTCTTCTCTAGTGAG	CATGTTTGGACTTCTGTGCTC
UBC	ATTTGGGTCGCGGTTCTTG	TGCCTTGACATTCTCGATGGT

Table 2.3. Primer sequences for qRT-PCR.

ER site	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
RARα	GCTGGGTCCTCTGGCTGTTC	CCGGGATAAAGCCACTCCAA
GREB1	GAAGGGCAGAGCTGATAACG	GACCCAGTTGCCACACTTTT
MYC	GCTCTGGGCACACACATTGG	GGCTCACCCTTGCTGATGCT
Negative control region	GCCACCAGCCTGCTTTCTGT	CGTGGATGGGTCCGAGAAAC

Table 2.4. Primer sequences for ChIP-qPCR at ER binding sites.

2.2 Methods

2.2.1 Mammalian cell culture methods

2.2.1.1 Thawing, growing and freezing cells

Cells were periodically cultured to between 80 and 90% confluence at 37°C with 5% CO_2, rinsed with sterile phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4), and trypsinised using 0.05% trypsin-EDTA. Trypsin was neutralised with DMEM or RPMI-1640 containing 10% FBS, and cells were centrifuged for 3 min at 1,000 g (Heraeus Megafuge, Thermo Fisher Scientific). The pelleted cells were

resuspended in media and replated depending on the initial confluence. To prepare frozen stocks, cells were resuspended in 10% DMSO and 90% FBS and frozen to -80°C using a Mr. Frosty (Nalgene). To revive frozen cell lines, stocks were thawed rapidly at 37°C, growth medium added and cells centrifuged for 3 min at 1,000 g before being resuspended in growth medium and placed into flasks.

2.2.1.2 Small interfering RNA transfections and drug treatments

Cells were grown in full medium for 24 hours and then transfected using Lipofectamine RNAiMAX transfection reagent according to manufacturer's instructions. Briefly, the transfection reagent was diluted in Opti-MEM (Gibco) and incubated for 5 minutes at room temperature. Subsequently, siRNA diluted in Opti-MEM was added to the transfection reagent/Opti-MEM mix and incubated for a further 20 minutes at room temperature. This mixture was added to cells in full growth medium and incubated for the relevant duration. For the purposes of obtaining protein, RNA or DNA samples, cells were washed twice in cold PBS and harvested in PBS containing protease inhibitors (Roche). For growth assays, cells were left in transfection medium for the duration of the assay. All siRNAs were used at a final concentration of 10 nM.

2.2.1.3 Analysis of cell growth

Cell growth after siRNA transfection was assessed using the IncuCyte® ZOOM Live Cell Analysis System (Essen BioScience). Cells were seeded in 96-well plates and transfected in at least quadruplicate as described in section 2.2.1.2, upon which plates were immediately placed in the IncuCyte® ZOOM Live Cell Analysis System (37°C with 5% CO_2) and growth was monitored for at least 120 hours via phase-contrast images taken every 3 hours. Confluence was assessed using default settings of the IncuCyte® ZOOM software.

2.2.2 RNA methods

2.2.2.1 RNA isolation and quantification

Cells were washed twice in cold PBS and harvested in PBS containing protease inhibitors (Roche). Total RNA was extracted using an RNeasy kit (Qiagen) according to

manufacturer's instructions, and quantified using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific).

2.2.2.2 Quantitative real-time PCR

Total RNA (1 µg) was used for cDNA conversion and qRT-PCR analysis. cDNA was synthesised using the SuperScript III Reverse Transcriptase kit (Invitrogen) according to manufacturer's instructions. For subsequent qPCR analysis, cDNA was diluted 1:10. Reactions were performed in triplicate, run on the BioRad CFX Connect RealTime System and analysed using BioRad CFX Maestro software version 1.1. Each reaction mixture contained 1X Power SYBR® Green PCR Master Mix (Applied Biosystems), forward and reverse primers (final concentration 10 µM each) and 2 µl diluted DNA template, with nuclease-free H_2O to a final volume of 15 µl. The hot-start *Taq* polymerase was heat-activated at 95°C for 10 minutes, followed by 45 cycles of 15 seconds at 95°C and 30 seconds at 60°C. The fluorescence was read in each cycle. A melt curve was constructed by increasing the temperature from 65 to 95°C and continuously reading the fluorescence. Expression relative to UBC as a housekeeping control was determined using the delta-delta Ct method (Livak and Schmittgen 2001).

2.2.2.3 RNA sequencing (RNA-seq)

2.2.2.3.1 Library preparation and sequencing

Library preparation and sequencing was carried out by the Genomics Core Facility (CRUK-CI). RNA libraries were prepared using the TruSeq stranded mRNA library prep kit (Illumina) and samples were sequenced on a HiSeq 4000 to approximately 30 million reads per sample.

2.2.2.3.2 RNA-seq data processing and bioinformatic analysis

RNA-seq data processing and bioinformatic analysis was performed by Dr Igor Chernukhin. 50 bp single-end reads were aligned to the Human Reference Genome (assembly hg38) using STAR version 2.5.2b (Dobin & Gingeras 2015). Read counts were normalised and tested for differential gene expression using DESeq2 (Love et al. 2014).

2.2.3 DNA methods

2.2.3.1 DNA isolation and quantification

Cells were washed twice in cold PBS and harvested in PBS containing protease inhibitors (Roche). DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer's instructions. DNA was RNase (Sigma) treated on-column (20 µl of 20 mg/mL) for 15 minutes at room temperature. Purified DNA was quantified using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific).

2.2.3.2 Mass spectrometry measurements of 5mC and 5hmC

Mass spectrometry analysis and data processing was performed by Dr Shiqing Mao (CRUK-CI). Cells were harvested and DNA isolated as described in section 2.2.3.1. Subsequently, 5mC and 5hmC were measured according to Bachman et al. (2014). Briefly, genomic DNA or ChIP DNA was incubated with 5 U DNA Degradase Plus (Zymo Research) for 4 hours at 37°C. Calibration curves were generated using a mixture of synthetic standards 2′-deoxycytidine (Sigma), 5-methyl- and 5-hydroxymethyl-2′-deoxycytidine (Berry & Associates), in the ranges 0.01–100 µM, 0.0005–5 µM and 0.0001–1 µM for C, 5mC and 5hmC, respectively. Samples and synthetic standards were spiked with an isotopically labelled mix that contained 1 µM 2′-deoxycytidine-($^{15}N,d_2$), 5-methyl-2′-deoxycytidine-(d_3) and 5-hydroxymethyl- 2′-deoxycytidine-(d_3) (Toronto Research Chemicals) as internal standards. Nucleotides were quantified using a Thermo Q-Exactive mass spectrometer. The quantitation was based on the peak area ratio of the analytes compared to their corresponding isotope-labelled internal standards, and the calibration curves. 5mC levels are expressed as a percentage of total cytosines.

2.2.3.3 Methyl Midi-seq (MMS)

2.2.3.3.1 DNA extraction, sample preparation and sequencing

Library preparation, sequencing and initial data processing for genomic analysis of 5mC was undertaken using the Methyl Midi-seq (MMS) service provided externally by Zymo Research (Irvine, California). Cells were harvested and DNA isolated as described in section 2.2.3.1. Subsequently, libraries were prepared from 500 ng of genomic DNA

sequentially digested with 60 U of TaqαI and 30 U of MspI (NEB) and then extracted with a DNA Clean & Concentrator kit (Zymo Research). Fragments were ligated to pre-annealed adapters containing 5'-methylcytosine instead of cytosine according to Illumina's specified guidelines (http://www.illumina.com). Adapter-ligated fragments of 150–250 bp and 250–350 bp were recovered from a 2.5% NuSieve 1:1 agarose gel using a Zymoclean Gel DNA Recovery Kit (Zymo Research). The fragments were then bisulfite-treated using the EZ DNA Methylation-Lightning Kit (Zymo Research). Preparative-scale PCR was performed and the resulting products were purified using a DNA Clean & Concentrator kit (Zymo Research) for sequencing on an Illumina HiSeq. Sequence reads from bisulfite-treated EpiQuest libraries were identified using standard Illumina basecalling software and then analysed using a Zymo Research proprietary analysis pipeline, which is written in Python and uses Bismark to perform the alignment (Krueger & Andrews 2011). Index files were constructed using the *bismark_genome_preparation* command and the entire reference genome. The *--non_directional* parameter was applied while running Bismark. All other parameters were set to default. Filled-in nucleotides were trimmed off when doing methylation calling. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T.

2.2.3.3.2 Methyl Midi-Seq bioinformatic analysis

Subsequent bioinformatic analysis was carried out by Dr Kamal Kishore (Bioinformatics Core Facility, CRUK-CI). The genome was partitioned into non-overlapping tiles of length 1 kb and 5mC was profiled within these tiles. Regions with low read coverage (less than 4 reads in any sample) were discarded from the analysis. The differential 5mC analysis was carried out by methylKit using Fisher Exact test (Akalin et al. 2012). For site-specific 5mC/5hmC analysis, shared ER/TET2 regions were defined using the *intersect* function in R to generate completely overlapping regions.

2.2.3.4 Reduced Representation Hydroxymethylation Profiling (RRHP)

DNA extraction, sample preparation and sequencing

Library preparation, sequencing and initial data processing for genomic analysis of 5hmC was undertaken using the Reduced Representation Hydroxymethylation Profiling (RRHP) service provided externally by Zymo Research (Irvine, California), as described in Petterson et al. (2014). Cells were harvested and DNA isolated as described in section

2.2.3.1. Subsequently, genomic DNA was fragmented overnight at 37°C with a hydroxymethyl-insensitive enzyme, MspI, and purified using the DNA Clean & Concentrator kit (Zymo Research). Modified Illumina TruSeq P5 and P7 adapters containing 5'-CG overhangs were ligated onto the digested DNA using T4 DNA ligase (2 hours at 16°C). Libraries were then strand extended at 72°C with Taq DNA Polymerase. The adapters were designed to regenerate the 5'-CCGG site at the P5 junction while the P7 adapter generates a 5'-TCGG junction, making it insensitive to MspI digestion. Adapterised libraries were treated with β-glucosyltransferase to label 5-hmC modifications and purified using a DNA Clean & Concentrator kit (Zymo Research). The glucosylated libraries were then subjected to an overnight MspI digestion at 37°C, cutting any fragments not containing a glucosyl-5hmC site at the P5 CCGG junction. After incubation, the libraries were size-selected from 100 bp to 500 bp and purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research). The fragments were amplified using OneTaq 2X Master Mix (NEB), with PCR conditions including an initial denaturation of 94°C for 30 seconds followed by 12 cycles of 94°C for 30 seconds, 58°C for 30 seconds, and 68°C for 1 minute. Fragments containing 5hmC were positively selected during PCR amplification with adapter-specific indexing primers whereas fragments lacking glucosylated-5hmC at the P5 junction were cleaved and, therefore, not amplified by PCR. Amplified libraries were purified using the DNA Clean and Concentrator kit, and multiplexed using equal volumes of the libraries. All adapters and primers used were synthesized by Integrated DNA Technologies. Sequence reads from RRHP libraries were first processed to trim off the low quality bases and the P7CG adapter at the 3' end of the reads. Reads were then aligned to the reference genome using bowtie (Langmead & Salzberg 2012) default parameters and the '–best' setting. Aligned reads with the MspI tag (CCGG) were counted. The correlation analysis between different RRHP libraries was performed by comparing the presence of the tagged reads at each profiled MspI site, and Pearson's coefficient was calculated accordingly.

2.2.3.4.1 Reduced Representation Hydroxymethylation Profiling bioinformatic analysis

Subsequent bioinformatic analysis was carried out by Dr Kamal Kishore (Bioinformatics Core Facility, CRUK-CI). The genome was partitioned into non-overlapping tiles of length 1 kb and 5hmC was profiled within these tiles. Regions with low read coverage (those with less than 10 reads in any sample) were discarded from the analysis. The differential 5hmC analysis was carried out using DESeq2 (Love et al. 2014). For site-specific 5mC/5hmC

analysis, shared ER/TET2 regions were defined using the *intersect* function in R to generate completely overlapping regions.

2.2.4 Protein methods

2.2.4.1 Western blot

Cells were harvested in RIPA buffer (Pierce) and the lysate sonicated using a Bioruptor Plus (Diagenode) for 2 minutes (30 seconds on/30 seconds off) to degrade the DNA. Protein was then quantified using a Direct Detect Spectrometer (Millipore). Samples were incubated at 70°C for 10 minutes in the presence of 1X LDS sample buffer (Invitrogen) and 1X NuPage sample reducing agent (Invitrogen) and loaded on NuPAGE 4-12% bis-Tris gels (Invitrogen). A Precision Plus Protein Standards molecular weight marker (BioRad) was used for determination of protein sizes. The gel tank apparatus was filled with 1X MOPS-SDS (3-(N-morpholino)propanesulfonic acid-sodium dodecyl suphate) running buffer (Invitrogen) and a voltage of 120 V applied for varying durations depending on the target protein size. Proteins were then transferred to nitrocellulose membranes using the iBlot® 2 Dry Transfer System (Invitrogen), or an overnight wet transfer (for TET2). The membrane was blocked using Odyssey Blocking Buffer (Li-Cor) for 1 hour at room temperature, and incubated with primary antibody overnight at 4°C. All primary antibodies (listed in Table 2.1) were used at a concentration of 1:1000, aside from the ERα antibody which was used at a concentration of 1:100. The membrane was then washed in Tris-buffered saline (50 mM Tris, 150 mM NaCl) containing 0.1% Tween-20 (TBS-T), incubated with fluorescent secondary antibodies (IRDye® 800 CW Goat anti-Mouse IgG 1:5,000 or IRDye® 680LT Goat anti-Rabbit 1:20,000, both Li-Cor) for 45 minutes at room temperature, and washed once more in TBS-T before imaging using the Odyssey CLx Imaging System (Li-Cor). Images were taken with the automated capture option of the Image Studio Version 4.0 software.

2.2.4.2 Parallel Reaction Monitoring (PRM)

2.2.4.2.1 Sample preparation and mass spectrometry

PRM sample preparation, method development and mass spectrometry (MS) analysis were performed by Dr Carmen Gonzalez Tejedo (Proteomics Core Facility, CRUK-CI).

Surrogate peptides unique to the target proteins of interest (ER, TET2, GATA3 and Actin) were chosen and stable-isotope-labelled versions of these peptides were synthesised as SpikeTides™ peptides by JPT Peptide Technologies, GmbH (Berlin, Germany). Cells were washed twice in cold PBS and harvested in PBS containing protease inhibitors (Roche). Cells were lysed and peptides digested with trypsin, and a mix of stable isotope-labelled peptide standards was added to the mixture. Mixtures were desalted using either Ultra-Micro C18 Spin Columns (Harvard Apparatus) or cartridges from an iST Sample Preparation Kit (Preomics) and reconstituted in either 3% acetonitrile/0.1% formic acid or the iST Sample Preparation Kit load buffer (Preomics). A Pierce Peptide Retention Time Calibration Mixture containing 15 synthetic heavy peptides mixed at an equimolar ratio (Thermo Scientific) was added to each sample at a final concentration of 20 fmol of peptides per 2 µg of total protein to assess chromatography performance and optimise scheduled MS acquisition windows. Diluted peptide mixtures were analysed by liquid chromatography mass spectrometry (LC-MS) on a Dionex Ultimate 3000 UHPLC system coupled to a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). Scheduled PRM transitions used a retention time window of 120 seconds. All samples were analysed in triplicate in the mass spectrometer.

2.2.4.2.2 PRM data processing and bioinformatic analysis

Data processing and bioinformatic analysis of PRM data was performed by Dr Carmen Gonzalez Tejedo (Proteomics Core Facility, CRUK-CI). All raw files were processed using Skyline-daily software v.19.0.9.190 (MacCoss Lab, University of Washington) for the generation of extracted-ion chromatograms and peak integration. Peak integrations were reviewed manually and transitions from analyte peptides were confirmed by the same retention times of the endogenous peptides and heavy stable isotope-labelled peptides time in a pre-selected retention time window. At least three transition ion peak area intensities were integrated and summed for each peptide (heavy and endogenous). The ratio of endogenous/heavy peak areas was calculated and the average of three independent injections of every sample was calculated to obtain a final quantification value for each peptide. Data were exported from Skyline for analysis and plotting using an in-house R script to calculate fold changes and p-values between different experimental conditions. Quantitative values obtained for actin peptides were used to normalise the data between different conditions.

2.2.4.3 Full proteome analysis

2.2.4.3.1 Sample preparation and mass spectrometry

Full proteome analysis was carried out by the Proteomics Core Facility (CRUK-CI), as described in Papachristou et al. (2018). Briefly, cells were washed twice in cold PBS and harvested in PBS containing protease inhibitors (Roche). Cells were sonicated in 200 µl of 0.1 M tetraethylammonium bromide (TEAB), 0.1% SDS (sodium dodecyl suphate) buffer followed by probe sonication and boiling at 95 °C. Protein concentration was estimated using a Bradford assay (BIO-RAD-Quick start) according to manufacturer's instructions. For each sample, 90 µg of total protein was reduced for 1 hour at 60°C by the addition of 2 µl 50 mM tris-2-carboxyethyl phosphine (TCEP, Sigma). Cysteines were blocked for 10 minutes at room temperature with the addition of 1 µl 200 mM methyl methanethiosulfonate (MMTS, Sigma). Proteins were digested overnight at 37°C in trypsin (Pierce) solution, added at a ratio of ~30:1 protein:trypsin. The following day peptides were labelled using the TMT (tandem mass tag) 11plex reagents (Thermo Scientific) with a randomised design. Peptides were fractionated on a Dionex Ultimate 3000 system at high pH using the X-Bridge C18 column (Waters) with 1% gradient. UV signal was recorded at 280 and 215 nm and fractions were collected in a peak-dependent manner. Peptide fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with a nano-ESI Fusion Lumos (Thermo Scientific).

2.2.4.3.2 Full proteome data processing and bioinformatic analysis

Data processing of full proteome results was carried out by the Proteomics Core Facility (CRUK-CI) according to Papachristou et al. (2018). Raw MS data was processed with the SequestHT search engine on the Proteome Discoverer 2.1 software for peptide and protein identifications. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20 ppm, Fragment Mass Tolerance 0.5 Da. Dynamic Modifications were Oxidation of M (+15.995 Da), Deamidation of N, Q (+0.984 Da) and Static Modifications were TMT6plex at any N-Terminus, K (+229.163 Da), Methylthio at C (+45.988). The Reporter Ion Quantifier node included a TMT 6plex (Thermo Scientific Instruments) Quantification Method. Further bioinformatic analysis was carried out by Dr Kamal Kishore (Bioinformatics Core Facility, CRUK-CI). Pre-processed quantitative datasets (peptide or protein-level intensities) generated by Proteome Discoverer were imported into R and data analysed using the qPLEXanalyzer tool (Papachristou et al.

2018), which uses analysis based on limma (an R/Bioconductor package) to identify differentially abundant proteins.

2.2.5 Chromatin Immunoprecipitation (ChIP)

2.2.5.1 Chromatin preparation and immunoprecipitation

ChIP was performed as described by Papachristou et al. (2018). Briefly, cells were crosslinked at room temperature by incubating with 2 mM disuccinimidyl glutarate (DSG) for 20 min followed by 1% formaldehyde for 10 min before crosslinking was quenched with 0.1 M glycine for 10 min. Cells were then washed twice in cold PBS and harvested in cold PBS containing protease inhibitors (Roche). Crosslinked cells were incubated with lysis buffer 1 (LB1, 50 mM Hepes–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40/Igepal CA-630, 0.25% Triton X-100) for 10 min followed by 5 min in LB2 (10 mM Tris–HCL, pH8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) before resuspending in LB3 (10 mM Tris–HCl, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na–Deoxycholate, 0.5% N-lauroylsarcosine). Chromatin was sonicated using the Bioruptor Plus (Diagenode) for 15 min (30 seconds on/30 seconds off) to generate DNA fragments of around 100-800 bp. Beads were pre-bound with antibody overnight at 4°C, with 5 µg of the appropriate antibody (or 2.5 µg each of ERα Abcam ab3575 and ERα Millipore 06-935 where these antibodies were used in a 1:1 combination) and 50 µl of Protein A Dynabeads (Invitrogen) was used for each immunoprecipitation. After washing to remove unbound antibody, chromatin and beads were combined and samples were immunoprecipitated overnight at 4°C. The following day, beads were washed 10 times in RIPA buffer (50mM HEPES pH 7.6, 1mM EDTA, 0.7% Na-deoxycholate, 1% NP-40, 0.5M LiCl) followed by 2 washes in Tris-EDTA (10mM Tris-HCl, 1mM EDTA). Chromatin was eluted and decrosslinked by incubating samples in elution buffer (50mM TrisHCl, pH8, 10mM EDTA, 1% SDS) for 6-18 hours at 65°C. Eluted DNA was treated with RNase A (20 ng/ml) for 1 hour followed by proteinase K (200 ng/ml) for 2 hours before DNA was purified by phenol-chloroform extraction and taken forward for either qPCR (section 2.2.5.2) or sequencing (section 2.2.5.3).

2.2.5.2 ChIP-qPCR

ChIP DNA was used neat and input DNA diluted 1:10 for qPCR analysis. Reactions were performed as described in section 2.2.2.2, using the primers listed in Table 2.4. Relative enrichment was determined as % of input.

2.2.5.3 ChIP-seq library preparation and sequencing

Library preparation was performed using the ThruPlex DNA-seq kit (Rubicon Genomics), and DNA was subjected to next generation sequencing on a HiSeq 4000 (Illumina) to reach approximately 30 million reads per sample. Sequencing was carried out by the Genomics Core Facility (CRUK-CI).

2.2.5.4 ChIP-seq data processing and bioinformatic analysis

ChIP-seq data processing and bioinformatic analysis was performed by Dr Igor Chernukhin. 50bp single-end reads were aligned to the Human Reference Genome (assembly hg38) using bowtie2 v.2.2.6 (Langmead & Salzberg 2012). Aligned reads with a mapping quality of less than 5 were filtered out. The read alignments from all replicates were combined into a single library and peaks were called using MACS2 version 2.0.10.20131216 (Zhang et al. 2008) with sequences from chromatin extracts from the same cell line or PDX used as a background input control. The peaks yielded with a MACS2 q-value ≤ $1e^{-3}$ were selected for downstream analysis.

2.2.5.4.1 Venn diagrams and differential binding analysis

For Venn diagrams, shared sites were defined as peaks overlapping by at least 1 bp, generated using bedtools v.2.29.0 (Quinlan & Hall 2010). Differential binding analysis was performed as described previously using DiffBind (Stark & Brown 2013). For heatmaps, MA plots, and average plots visualising tag density and signal distribution, consensus peak sets across the compared conditions were determined using DiffBind. Heatmaps and average plots were generated with the read coverage in a window of ± 5 kb flanking the tag midpoint using a bin size of 1/100 of the window length. MA plots were generated in R v.3.5.1 or later, and average plots, boxplots and heatmaps were generated using MATLAB.

2.2.5.4.2 Motif analysis

MEME Suite (v.4.9.1) tools were used for motif analysis. FIMO (Grant et al. 2011) was used to search all known transcription factor motifs from the JASPAR database (JASPAR CORE 2016 vertebrates) in tag-enriched sequences. Peak size-matched, randomly selected open chromatin regions based on an MCF7 MNase dataset (EBI Array Express E-MTAB-1958) were used as background controls. The motif frequencies for both tag-enriched and control sequences were calculated as the sum of motif occurrences adjusted with MEME q-value. Motif enrichment analysis was performed by calculating the odds of finding an overrepresented motif among MACS2-defined peaks by fitting Student's t-cumulative distribution to the ratios of motif frequencies between tag-enriched and background sequences. Yielded p-values were further adjusted using Benjamini-Hochberg correction. MEME (Bailey & Elkan 1994) and DREME (Bailey 2011) were used to perform *de novo* motif analysis on sequences corresponding to ChIP-seq peak regions and the resulting position weight matrix was compared to the JASPAR, Transfac and UNIPROBE databases by the TOMTOM application (Gupta et al. 2007). A p-value 0.0001 was used as a threshold to define the presence of a motif.

2.2.6 Rapid Immunoprecipitation Mass-spectrometry of Endogenous Proteins (RIME)

2.2.6.1 Chromatin preparation and immunoprecipitation

Chromatin preparation and immunoprecipitation was performed as described in section 2.2.5.1, until the bead washing step after overnight immunoprecipitation. At this point, beads were washed 10 times in RIPA buffer (50mM HEPES pH 7.6, 1mM EDTA, 0.7% Na-deoxycholate, 1% NP-40, 0.5M LiCl) followed by two washes in 100 mM ammonium hydrogen carbonate (AMBIC). Washed beads were frozen at -20°C prior to peptide digestion and RIME analysis (section 2.2.6.2).

2.2.6.2 RIME sample preparation and mass spectrometry

RIME sample preparation and analysis was performed by the Proteomics Core Facility (CRUK-CI) as described in Glont et al. (2019) and Papachristou et al. (2018). Briefly, tryptic digestion of bead-bound protein was performed by addition of 10 µl trypsin solution (15 ng/µl) (Pierce) in 100 mM AMBIC followed by overnight incubation at 37 °C. A second

digestion step was performed the next day for 4 h. After tryptic digestion tubes were placed on a magnetic stand to allow removal of the supernatant after acidification by the addition of 2 µl 5% formic acid. Digested peptide mixes were cleaned using Ultra-Micro C18 Spin Columns (Harvard Apparatus) according to manufacturer's instructions. For non-quantitative RIME, digested peptide mixtures were diluted 1/10 with loading buffer (0.1% formic acid, 2% acetonitrile (ACN), water) and analysed on an LTQ Velos-Orbitrap MS (Thermo Scientific) coupled to an Ultimate RSLCnano-LC system (Dionex). For quantitative RIME (qPLEX-RIME), digested, cleaned peptide samples were dried using a speedvac, reconstituted in 100 µl 0.1 M TEAB and labelled using TMT 10plex reagents (Thermo Fisher) with a randomised design. The peptide mixture was fractionated with Reversed-Phase cartridges at high pH (Pierce). Nine fractions were collected using different elution solutions in the range of 5–50% ACN. Peptide fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with a nano-ESI Fusion Lumos (Thermo Scientific).

2.2.6.3 RIME data processing and bioinformatic analysis

2.2.6.3.1 Non-quantitative RIME

Data processing of RIME results was carried out by the Proteomics Core Facility (CRUK-CI) according to Papachristou et al. (2018). The raw MS files were processed with the SequestHT search engine on Proteome Discoverer 2.1 software for peptide and protein identifications. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20ppm, Maximum Missed Cleavages sites 2, Fragment Mass Tolerance 0.02Da. Dynamic Modifications were Oxidation of M (+15.995Da) and Deamidation of N, Q (+0.984Da). The coverage plots were created using the qPLEXanalyzer tool (Papachristou et al. 2018). Further bioinformatic analysis was carried out by Dr Kamal Kishore (Bioinformatics Core Facility, CRUK-CI) according to Papachristou et al. (2018). Specific interactors were considered as those occurring in at least two out of three independent replicates. Any proteins that appeared in any one of the three IgG control RIME experiments were excluded.

2.2.6.3.2 Quantitative RIME

Data processing of RIME results was carried out by the Proteomics Core Facility (CRUK-CI) according to Papachristou et al. (2018). The raw MS files were processed with the SequestHT search engine on the Proteome Discoverer 2.1 software for peptide and protein identifications. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20 ppm, Fragment Mass Tolerance 0.5 Da, Dynamic Modifications were Oxidation of M (+15.995 Da), Deamidation of N, Q (+0.984 Da) and Static Modifications were TMT6plex at any N-Terminus, K (+229.163 Da). The Reporter Ion Quantifier node included a TMT 6plex (Thermo Scientific Instruments) Quantification Method. Further bioinformatic analysis was carried out by Dr Kamal Kishore (Bioinformatics Core Facility, CRUK-CI). Pre-processed quantitative datasets (peptide or protein-level intensities) generated by Proteome Discoverer were imported into R and data analysed using the qPLEXanalyzer tool (Papachristou et al. 2018), which uses analysis based on limma (an R/Bioconductor package) to identify differentially abundant proteins.

2.2.7 Survival analysis

For analysis of relapse free survival, Kaplan-Meier plotter (http://kmplot.com) was used. The data and methods used for the analysis are described in Győrffy et al. (2010). Briefly, patients were stratified into high or low expression groups according to the median level of the gene probe selected (TET2 JetSet probe 227624_at). ER+ and ER- cohorts were analysed separately, with the “ER status derived from GE data” option selected. All other parameters were left as default.

2.2.8 Additional software and statistical tests

All analysis additional to that described in the method-specific sections above was performed using either MS Excel, GraphPad Prism v.8 or R v.3.5.1. Significance was assessed using Student’s t-test or Welch’s t-test. Only values with a p-value less than 0.05 were considered statistically significant. Error bars represent standard deviation (SD).

Chapter 3

The effect of GATA3 depletion on the ER complex, and the identification of TET2 as a key ER interactor

3.1 Introduction

ER drives tumour development in ER+ breast cancer in co-operation with numerous other proteins that together form the ER transcriptional complex. Although there are many key players that participate in the ER complex whose function is well understood, the roles of many other ER cofactors remain to be fully characterised, and it is likely there are yet further ER co-operating proteins yet to be found. These studies are important for developing a clearer picture of the molecular mechanisms driving the development and progression ER+ breast cancer, towards the aim of more effective targeted therapies.

GATA3 is a transcription factor fundamental to mammary gland development that is highly expressed, and frequently mutated, in ER+ breast cancers (Cancer Genome Atlas Network 2012; Kouros-Mehr et al. 2006; Perou et al. 2000; Sorlie et al. 2001). Oestrogen-induced growth of ER+ breast cancer cells appears dependent on GATA3 (Eeckhoute et al. 2007; Kong et al. 2014), and at the molecular level, GATA3 motifs are enriched around ER binding sites (Carroll et al. 2005; Lin et al. 2007; Serandour et al. 2013). ER and GATA3 co-localise at a large proportion (~45%) of ER binding sites in ER+ breast cancer cells (Kong et al. 2011; Theodorou et al. 2013), therefore it appears that there may be direct functional interplay between these two proteins that contributes to the breast cancer phenotype.

In terms of the mechanisms behind this potential co-operativity, GATA factors have been shown to interact directly with nucleosomes *in vitro* (Cirillo et al. 2002; Takaku et al. 2016) implying pioneer factor-like activities, as have been shown for FOXA1 in the context of ER. However, although overexpression of GATA3 in MBA-MD-231 cells demonstrates GATA3-induced chromatin remodelling at a subset of sites, rather than appearing an activity intrinsic to GATA3, this instead correlates with its capacity to recruit ATP-dependent chromatin remodellers such as BRG1 (Takaku et al. 2016). Moreover, silencing of GATA3 in ER+ breast cancer cells affects the distribution of ER binding, but with equal proportions of both stronger and weaker ER binding events observed (Theodorou et al. 2013). Nevertheless, these effects are associated with oestrogen-independent changes in FOXA1, p300, H3K4me1 and H3K27ac distribution that correlate with gene expression changes upon subsequent oestrogen stimulation, implying that the importance of GATA3 in ER biology may be partly due to a role in regulating enhancer accessibility. In the same study, GATA3 loss was further implicated in ER signalling through the examination of chromatin loops at the promoters of the oestrogen-regulated genes *TFF1* and *TFF3*. At these sites, loss of GATA3 resulted in changes in chromatin looping that were associated with increased expression of these ER-regulated genes. This paints a complex picture of the role of GATA3 in ER biology, where, despite their putative functional connection and similar developmental roles, GATA3 may not always reinforce ER activity, but may instead fine-tune its effects on transcription. Importantly, whilst these observations of the contribution of GATA3 to ER signalling are themselves complex, the potential mechanisms behind them remain unclear. In particular, given that GATA3 appears to lack intrinsic enzymatic or chromatin remodelling capabilities, investigating the proteins that are recruited to the ER complex by GATA3 could shed further light on its role.

Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) is a method that has previously been used to study the composition of the ER complex, and to identify novel ER-interacting proteins (Mohammed & Carroll 2013). This technique involves antibody-based purification of endogenous protein complexes, with formaldehyde and disuccinimidyl glutarate (DSG) crosslinking performed prior to immunoprecipitation to enhance the scope of interactions that can be detected. A quantitative version of RIME (quantitative multiplexed rapid immunoprecipitation mass spectrometry of endogenous proteins, qPLEX-RIME) involves tandem mass tag (TMT)

labelling of samples prior to peptide fractionation and subsequent mass spectrometry. This facilitates precise quantitation of changes in chromatin-associated protein interactomes between different experimental conditions. For example, this technique has recently been used to investigate changes in the composition of the ER complex in response to tamoxifen treatment (Papachristou et al. 2018). To gain insights into the contribution of GATA3 to ER signalling, qPLEX-RIME was employed to evaluate the ER complex under control and GATA3-depleted conditions. The aim of this approach was to either detect mediators of GATA3 action that may be lost as a result of its absence from the ER complex, or to identify compensatory factors that could provide an indication of GATA3 activity, to gain further insights the potential contribution of this protein to ER-regulated transcription.

3.2 Results

3.2.1 GATA3 knockdown affects the composition of the ER complex

GATA3 was robustly depleted in MCF7 cells after 48 hours of siRNA-mediated knockdown, with no effect on total ER protein levels, as confirmed by Western Blot (Figure 3.1A). ER qPLEX-RIME was then conducted comparing the ER interactome under control or GATA3-silenced conditions. As expected, GATA3 was observed as the most significantly depleted ER interactor, validating the knockdown approach (Figure 3.1B). Beyond this, and notably, given the implied functional link between GATA3 and ER, only a small number of significant changes were observed in response to GATA3 knockdown. The proteins significantly enriched in the ER complex in response to loss of GATA3 were the transcription factors LIM-homeobox 4 (LHX4) and zinc finger and BTB domain containing protein 34 (ZBTB34). The protein GREB1L (growth regulation by estrogen in breast cancer 1-like protein), which is of unknown function but has been implicated as a co-activator for the retinoic acid receptor (RAR) (Brophy et al. 2017; Herlin et al. 2019), was also enriched. Concurrent with this, the only protein that was significantly depleted in response to GATA3 knockdown (other than GATA3 itself) was the dioxygenase enzyme TET2. RNA-seq demonstrated that these changes to the ER complex in response to GATA3 knockdown are likely due to altered expression of these proteins (Figure 3.2), suggesting that the decreased TET2-ER interactions might result from downregulation of

TET2 levels in response to GATA3 loss, and the increased presence of LHX4, ZBTB34 and GREB1L may be linked to their protein-level upregulation. Interestingly, it has been suggested that both LHX4, and the BTB/POZ protein family (of which ZBTB34 is a member) may be capable of binding to methylated DNA (Filion et al. 2006; Qi et al. 2006; Yin et al. 2017). TET2 also has a link to DNA methylation through its capacity to regulate DNA modifications, mediating the conversion of 5mC to 5hmC, followed by iterative oxidation of 5hmC to 5fC and 5caC (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009). These results suggest a potential role for GATA3 in modulating reading and writing of DNA modifications as part of the ER complex.

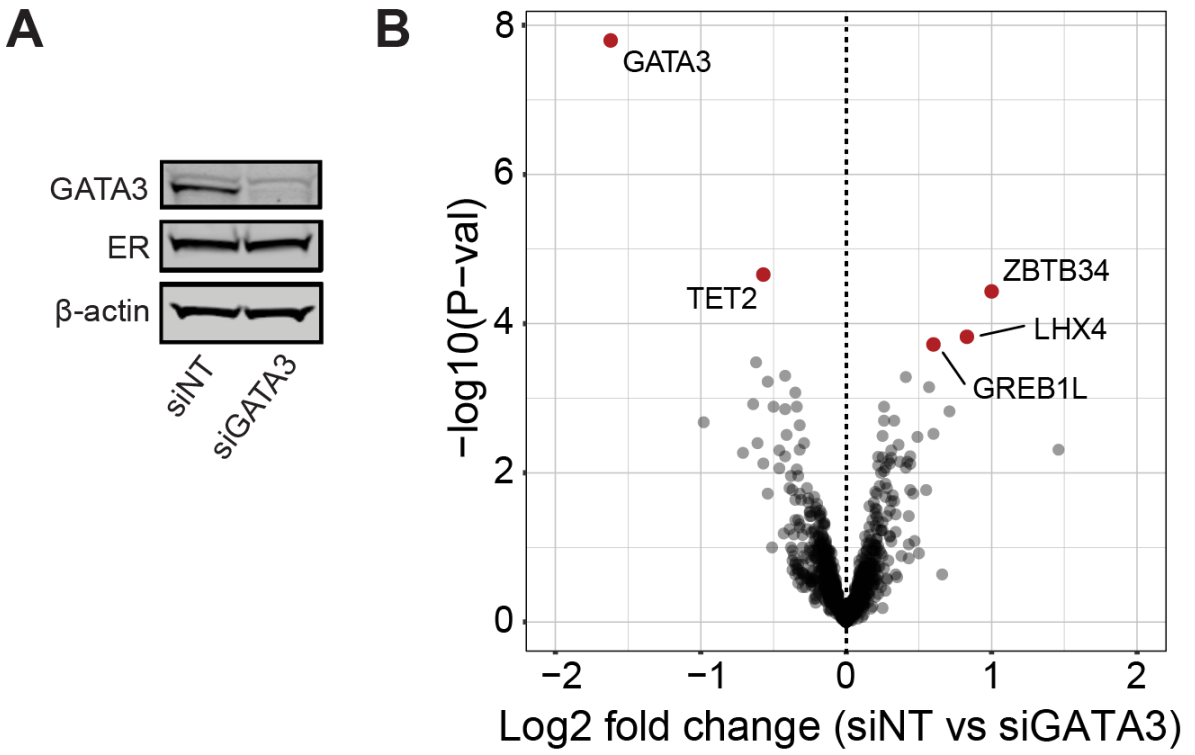


Figure 3.1. GATA3 knockdown alters the composition of the ER complex.

A) Western blot for GATA3 and ER after 48 hours' treatment with either non-targeting control siRNA (siNT), or siRNA targeting GATA3 (siGATA3), demonstrating robust GATA3 depletion with no effect on total ER levels. β-actin is used as a loading control. B) ER qPLEX-RIME in MCF7 cells showing changes to the ER complex after GATA3 knockdown (48 hours). Four replicates of ER RIME and one pooled IgG control RIME for each condition were included in each 10plex TMT MS run. Significantly enriched or depleted proteins according to the adjusted p-value are highlighted in red (p adj ≤ 0.05 after multiple testing correction using Benjamini-Hochberg procedure).

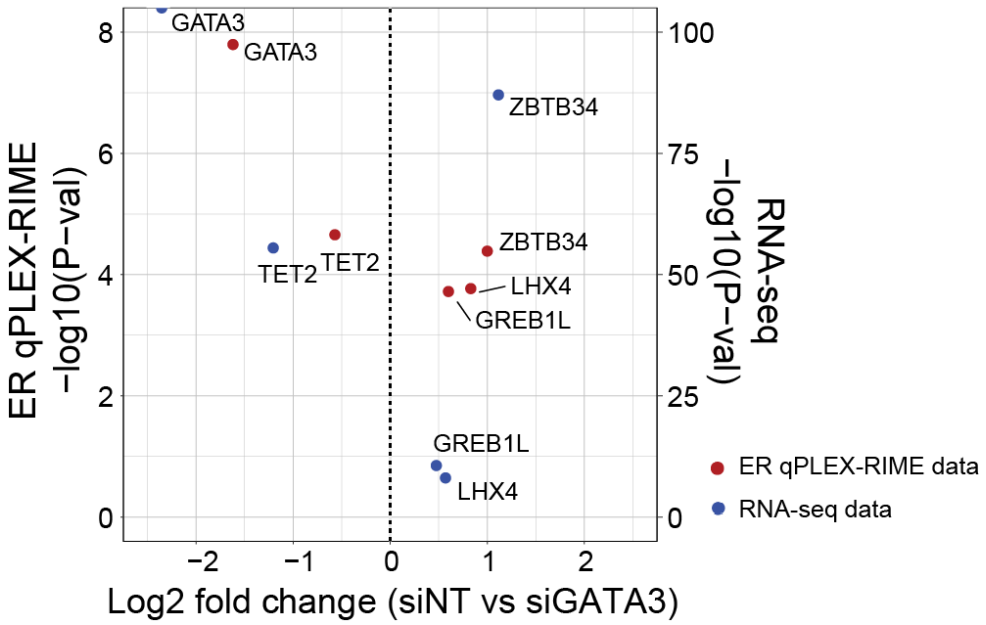


Figure 3.2. Protein-level changes in the ER complex correspond to gene expression changes induced by GATA3 knockdown.

For the five significantly changing proteins identified in Figure 3.1, significance in ER qPLEX-RIME (p-value, left y axis, red data points) and RNA-seq (p-value, right y axis, blue data points) were plotted against log2 fold change, showing that changes in the expression levels of these genes in response to GATA3 knockdown likely contribute to their altered levels in the ER complex.

3.2.2 TET2 and GATA3 antibody testing using RIME

Given the capacity of TET2 for regulating DNA modifications, and the developing focus on the role of this factor in transcriptional regulation (Chen et al. 2018; Li et al. 2017; Rasmussen et al. 2015, 2019; Wang et al. 2018; Wang et al. 2015), its loss from the ER complex in response to GATA3 depletion warranted further investigation. Due to the lack of robust, commercially available IP-grade antibodies for TET2, several groups have circumvented the resultant challenges associated with mapping endogenous TET2-chormatin interactions through generating custom TET2 antibodies, or by overexpressing TET2 tagged with additional epitopes more easily detectable using ChIP (Rasmussen et al. 2019; Wang et al. 2015; Wang et al. 2018). RIME provides an unbiased and sensitive approach for assessing the specificity and efficacy of antibodies for immunoprecipitation, so for this study, a selection of commercially available TET2 antibodies were tested using this method. To allow assessment of the reciprocal interaction between TET2 and GATA3, antibody candidates against GATA3 were also evaluated.

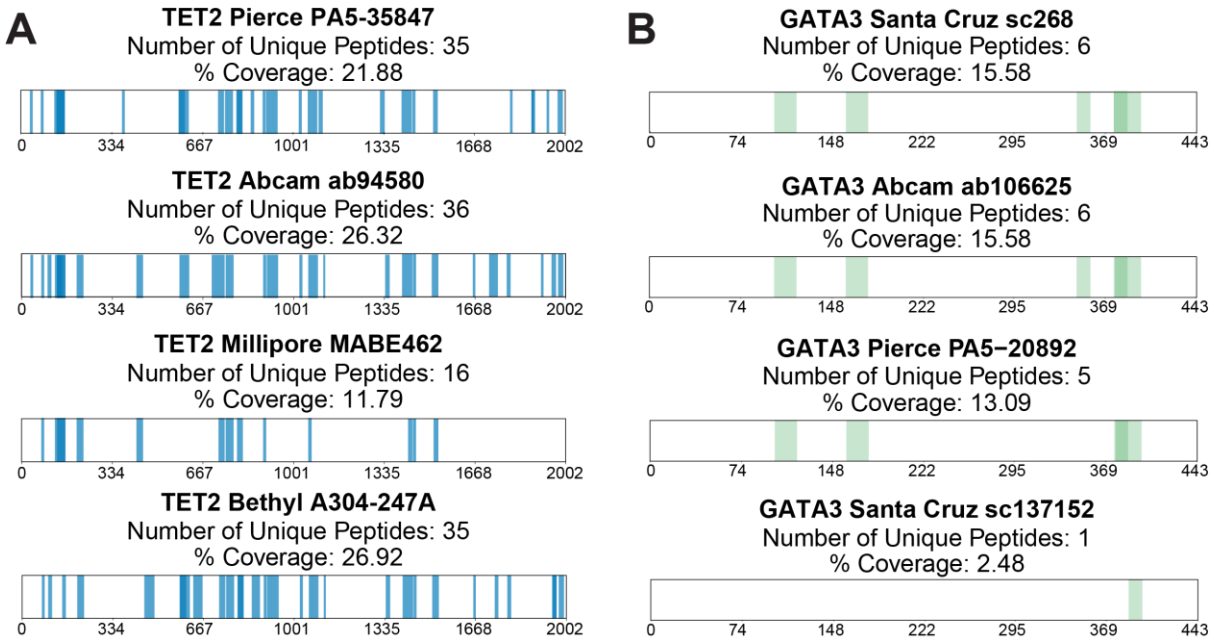


Figure 3.3. Coverage plots for TET2 and GATA3 antibodies tested using RIME. To assess the efficiency of each antibody to immunoprecipitate its target protein, plots showing the peptide sequence coverage of the bait protein were generated for A) TET2 and B) GATA3. Plots show the location of unique peptides identified with high confidence across the protein sequence. The corresponding percentage coverage and number of unique peptides is detailed above each plot. Data are the results of one replicate, used as a quick way of assessing multiple antibodies prior to more detailed experimental work.

Non-quantitative RIME was performed on untreated, asynchronous MCF7 cells. Antibodies were assessed through measuring protein sequence coverage in terms of the unique peptides detected by mass spectrometry. RIME using species-matched IgG antibodies as negative controls was performed in parallel to discriminate specific from non-specific interactions. As shown in Figure 3.3A, the TET2 protein was immunoprecipitated with varying efficiencies using the different antibody candidates. Given that cell lysates are crosslinked for RIME, this antibody validation also provided the potential to explore TET2-associated proteins. The two antibodies achieving the highest TET2 sequence coverage were Abcam ab94580 and Bethyl A304-247A (Figure 3.3A), so to assess the consistency of the TET2 interactors detected using these two reagents, the lists of proteins detected in each of these RIME experiments were overlapped with one another, as shown in Figure 3.4A. Proteins detected in the IgG negative control are also illustrated in the Venn diagram to provide an indication of antibody specificity. Both antibodies resulted in the detection of similar numbers of proteins (648 interactors using

Abcam ab94580 and 587 interactors using Bethyl A304-247A), with 396 of these proteins common to both antibodies. Of these, 97 (~25%) were also detected in the IgG negative control, leaving 299 of these shared interactors as specific to TET2. Bethyl A304-247A and Abcam ab94580 antibodies thus appeared similar both in their ability to purify TET2, and in terms of the interactors co-purified, and Abcam antibody ab94580 was selected for use in future work.

As shown in Figure 3.3B, GATA3 was also immunoprecipitated with varying efficiencies using the four different antibodies tested against this protein. The two antibodies achieving the highest sequence coverage were Santa Cruz sc268 and Abcam ab106625, both resulting in detection of 15.58% of the GATA3 protein. To compare the GATA3 interactors co-purified using each of these antibodies, the lists of detected interactors were overlapped, shown in Figure 3.4B. Proteins detected in the IgG negative control are also illustrated in the Venn diagram. Some consistency was observed between the interactors detected using these antibodies, however the number of interactors detected using the Santa Cruz antibody was much smaller (110 interactors, versus 525 for Abcam ab106625). One explanation for this could be that the polyclonal Abcam ab106625 is less susceptible to epitope masking than the monoclonal Santa Cruz sc268, resulting in a more robust pull-down and facilitating the detection of a greater number of interactors. The overlap of these two sets of interactors with those detected in the IgG negative control illustrates their specificity (Figure 3.4B). A proportion of the interactors detected using both Santa Cruz sc268 and Abcam ab106625 appeared to be non-specific based on their co-occurrence in the IgG negative control, however non-specific interactors constituted a much larger proportion of the total Santa Cruz sc268 list (73 out of 110 proteins, ~66%) than the Abcam ab106625 list (124 out of 525 proteins, ~24%). The increased scope of Abcam ab106625 for detecting specific GATA3 interacting proteins thus prompted the selection of this reagent for GATA3 RIME.

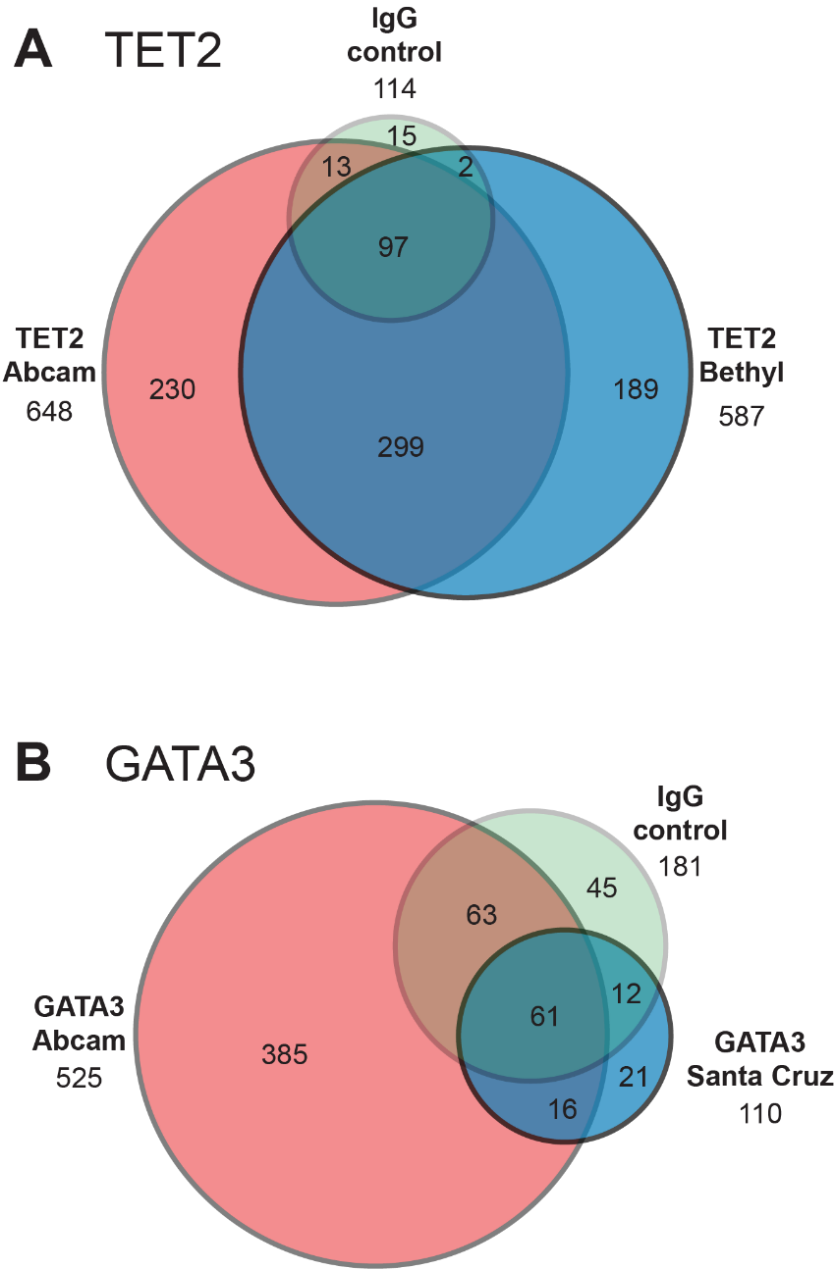


Figure 3.4. Venn diagrams of TET2 and GATA3 interactors detected using different antibodies in RIME.

Lists of Uniprot IDs of all interactors identified using selected A) TET2 and B) GATA3 antibodies were overlapped. The proportion of unique interactors detected by each antibody is indicated by overlap with lists of interactors detected using species-matched IgG negative controls. Antibody species: TET2 Abcam ab94580 (rabbit); TET2 Bethyl A304-247A (rabbit); GATA3 Abcam ab106625 (rabbit); GATA3 Santa Cruz sc268 (mouse). For the GATA3 Venn diagram, a combined list of proteins detected in both rabbit and mouse IgG RIME was used.

3.2.3 ER, GATA3 and TET2 form a complex, and share common interacting proteins

Using the antibodies selected for GATA3 and TET2 RIME in section 3.2.2, and the antibody combination previously optimised for ER immunoprecipitation by Glont et al. (2019), RIME for ER, GATA3 and TET2 was performed in MCF7 cells. Biological triplicates were performed for each factor, including an IgG negative control. Specific interactors for each protein were defined as those detected in at least two out of three independent replicates, and absent from all three replicates of the IgG negative control. ER, GATA3 and TET2 were reciprocally detected as interactors of one another, implying that these proteins interact closely at the core of the ER complex. In addition, ER, GATA3 and TET2 demonstrated a large number of common interactors (379 proteins). The Venn diagram in Figure 3.5B demonstrates the overlap of the specific interactors detected for ER, GATA3 and TET2, whilst Figure 3.6 shows the full list of interactors common to all three proteins. ER, GATA3 and TET2 common interactors included several other key ER-associated proteins such as FOXA1, GREB1, RARα, and ER co-activators NCOA3 and CARM1 (Anzick et al. 1997; Chen et al. 2000; Deschênes et al. 2007; Mohammed et al. 2013; Ross-Innes et al. 2010), demonstrating that TET2 interacts with central components of the ER machinery. The sequence coverage achieved for ER, GATA3 and TET2 in these experiments is shown in Figure 3.5A, demonstrating that each bait protein was robustly detected using RIME. Additionally, the sequence coverage and corresponding number of unique peptides obtained for ER, GATA3 and TET2 in all three RIME experiments are shown in Figure 3.7.

ER RIME performed in this manner has been described previously, and the list of ER interactors identified in this experiment matched up well with existing results (Glont et al. 2019; Papachristou et al. 2018). Though GATA3 RIME has not been published, GATA3 is consistently identified as an ER interactor in published ER RIME datasets in MCF7 cells (Mohammed et al. 2016; Glont et al. 2019; Papachristou et al. 2018), thus the detection of many key ER interactors in the GATA3 RIME validates the GATA3 antibody used for this approach. Additionally, amongst the proteins identified in the GATA3 RIME was the known GATA3-interacting protein Friend of GATA1 (FOG1), also known as ZFPM1 (Cantor & Orkin 2005), further confirming that this antibody captures direct interactors of GATA3.

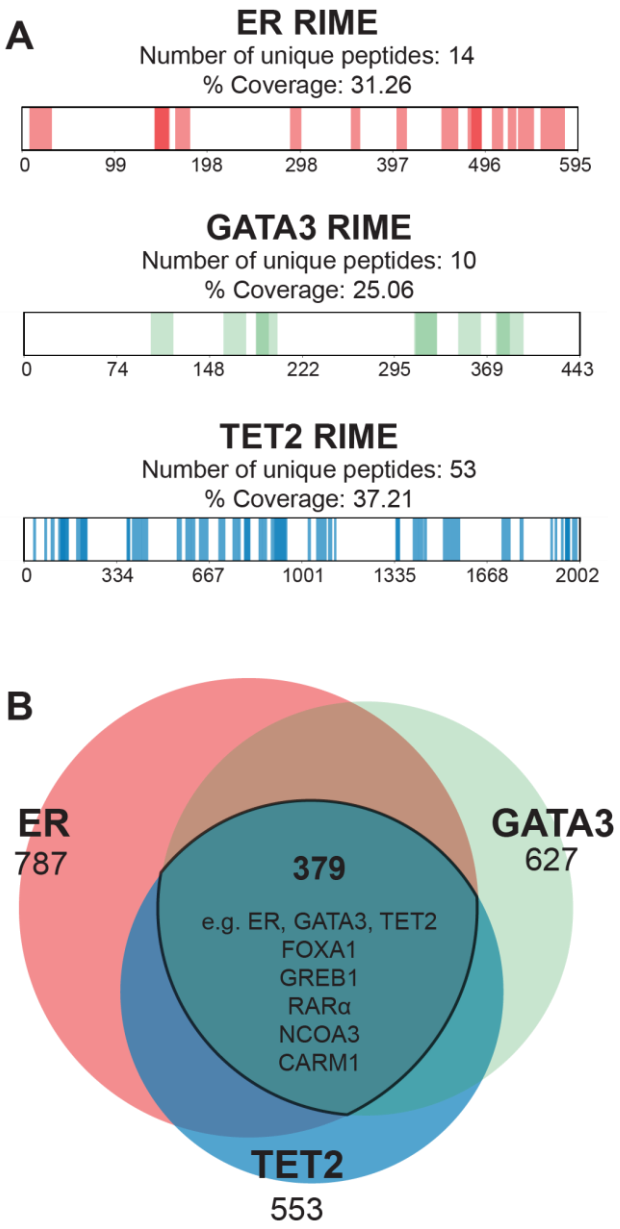


Figure 3.5. ER, GATA3 and TET2 form a complex, and share common interacting proteins.

A) Non-quantitative RIME in MCF7 cells yielded robust coverage of the three bait proteins ER, GATA3 and TET2, indicated by coverage diagrams showing the unique peptides identified with high confidence across the protein sequence. The number of unique peptides and corresponding % sequence coverage is shown above each plot. Each diagram provides a representative example of three biological replicates. B) Overlap of lists of Uniprot IDs of specific interactors for ER, GATA3 and TET2. Specific interactors were defined as those occurring in at least two out of three independent replicates. Proteins that appeared in any one of the three IgG control RIME experiments were excluded. ER, GATA3 and TET2 shared a total of 379 common interactors by RIME. Several key ER complex proteins were amongst these, highlighted in the central portion of the diagram. ER, GATA3 and TET2 common interactors are listed in full in Figure 3.6.

ACIN1	DAZAP1	GATA3	LUC7L	PRDX1	RPL17	SF3B4	TKT
ACP1	DBR1	GATAD2B	LUC7L2	PRKDC	RPL21	SLC9A3R1	TLE1
ACTC1	DCAF7	GMPS	LUC7L3	PRMT1	RPL22	SLTM	TLE3
ADAR	DDB1	GRB2	MATR3	PRMT6	RPL24	SMARCA4	TLK2
AES	DDX1	GREB1	MBNL2	PRPF19	RPL27	SMARCB1	TMPO
AHCY	DDX17	GRHL2	MCM3	PRPF3	RPL3	SMARCC1	TNPO1
AHDC1	DDX21	GTF2I	MCM7	PRPF31	RPL34	SMARCC2	TPM4
AKAP8	DDX23	H1F0	MEMO1	PRPF38B	RPL35	SMARCD2	TPR
AKAP8L	DDX39B	H2AFY	MFAP1	PRPF4	RPL35A	SMARCE1	TRA2A
API5	DDX3X	H2AFZ	MLF2	PRPF6	RPL4	SMC1A	TRA2B
ARF3	DDX42	HDAC2	MOV10	PRPF8	RPL6	SMU1	TRIM24
ARGLU1	DDX46	HIST1H2AB	MTA2	PSPC1	RPL7	SNRNP200	TRIM25
ARID1A	DDX5	HIST1H2BJ	MYEF2	PTBP1	RPRD2	SNRNP40	TRIM28
ARID1B	DDX6	HIST1H3A	MYH14	PTBP3	RPS10	SNRPA1	TRIM33
ARL6IP4	DHX15	HMCES	MYH9	PUF60	RPS11	SNRPB	TRIP6
BCL9L	DKC1	HMGB2	MYL6	QKI	RPS13	SNRPD1	TRPS1
BCLAF1	DNAJA1	HMGB3	NACC1	RACK1	RPS14	SNRPD2	TUBB
BOLA2	DNAJB1	HNRNPA0	NAT10	RAI1	RPS15	SNRPE	U2AF1
BUB3	DYNLL1	HNRNPAB	NCBP1	RALY	RPS15A	SNW1	U2AF2
CARM1	EEF1D	HNRNPC	NCBP2	RAN	RPS16	SRRM1	U2SURP
CBX3	EEF2	HNRNPD	NCOA3	RARA	RPS2	SRRM2	UBAP2L
CBX5	EFTUD2	HNRNPDL	NCOA5	RBBP7	RPS20	SRRT	UBC
CCAR1	EIF3I	HNRNPF	NCOR1	RBM10	RPS21	SRSF1	UGDH
CCAR2	EIF4A1	HNRNPH2	NCOR2	RBM12	RPS24	SRSF10	UPF1
CDC5L	EIF4A3	HNRNPH3	NKRF	RBM12B	RPS25	SRSF5	USP39
CEBPB	EIF4B	HNRNPL	NMD3	RBM14	RPS27	SRSF6	XPO1
CELF1	EIF4H	HNRNPLL	NONO	RBM15	RPS28	SRSF7	XRCC5
CFL1	EIF5A	HNRNPM	NOP58	RBM22	RPS3A	SRSF9	XRCC6
CHERP	ELAVL1	HNRNPR	NR2F2	RBM25	RPS4X	SSB	XRN2
CIRBP	ENO1	HNRNPU	NRIP1	RBM3	RPS7	SSRP1	YLPM1
CLIC3	ERH	HNRNPUL1	NUDT16L1	RBM33	RPS9	STARD10	YTHDF2
CLK3	ESR1	HNRNPUL2	NUDT21	RBM39	RSRC2	STAT3	YWHAQ
CLTC	ESRP1	HSP90AA1	NUMA1	RBM4	RTCB	STAU1	ZC3H11A
CPNE3	ESRP2	HSPA1B	NXF1	RBM45	RTFDC1	STRBP	ZC3H14
CPSF1	FAM120A	HSPA8	OGT	RBM4B	RXRA	SUB1	ZC3H18
CPSF6	FAM50A	ILF2	PA2G4	RBM8A	S100A11	SUMO1P1	ZC3H4
CPSF7	FAM98B	KDM1A	PABPC1	RBMX	SAFB	SUPT16H	ZFR
CREBBP	FASN	KHDRBS1	PABPN1	RBPMS	SAFB2	SUPT5H	ZNF207
CRIP2	FEN1	KHSRP	PARP1	RCC1	SARNP	SYNCRIP	ZNF217
CRNKL1	FIP1L1	KPNA2	PCBP1	RCC2	SART1	TAF15	ZNF281
CSDE1	FOSL2	KRT14	PCBP2	RNF20	SART3	TAGLN2	ZNF326
CSE1L	FOXA1	L1RE1	PDCD6	RNF40	SBNO2	TBL1XR1	ZNF385A
CSNK1A1	FUBP1	LARP4	PFN1	RNH1	SET	TBX2	ZNF638
CSTF1	FUBP3	LASP1	PFN2	RPL10	SF1	TCERG1	
CSTF2	FUS	LGALS3	PIAS3	RPL13	SF3A3	TCF20	
CSTF3	FXR1	LMNA	PKM	RPL13A	SF3B1	TET2	
CTBP1	G3BP1	LMNB1	POLDIP3	RPL14	SF3B2	THRAP3	
CTBP2	GAPDH	LMX1B	PPIA	RPL15	SF3B3	TIAL1	

Figure 3.6. Full list of ER, GATA3 and TET2 common interactors from RIME in MCF7 cells.

Specific interactors were considered as those occurring in at least two out of three independent replicates. Any proteins that appeared in any one of three IgG negative control RIME experiments were excluded.

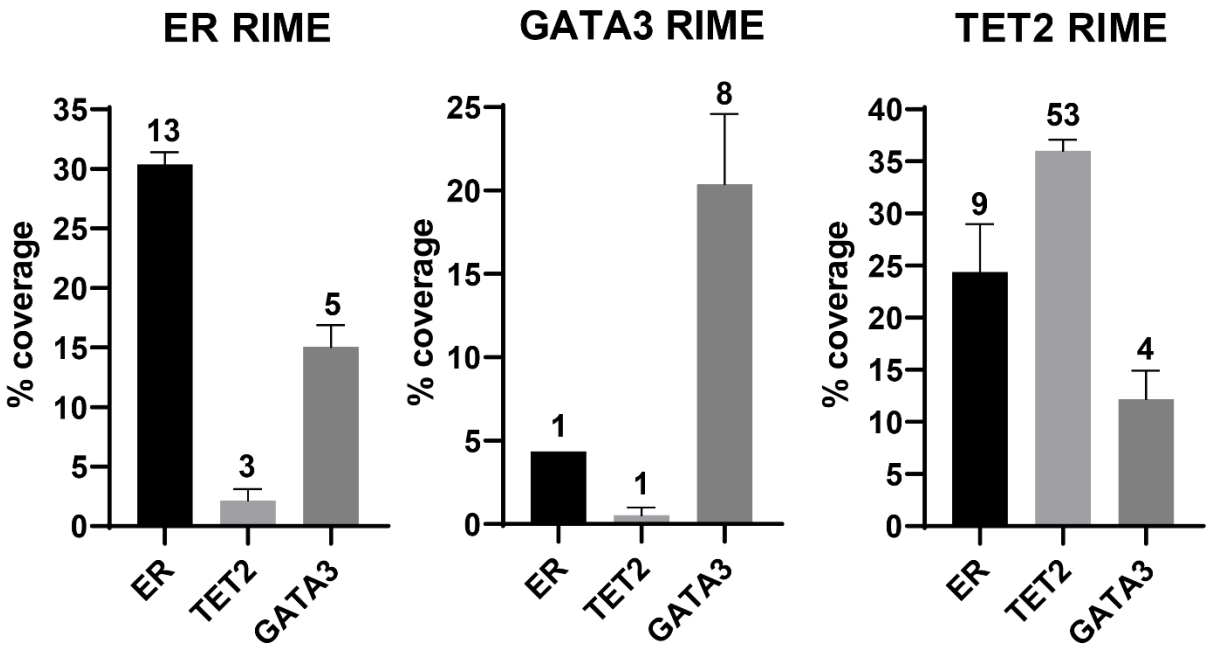


Figure 3.7. Protein coverage and unique peptides for reciprocal detection of ER, GATA3 and TET2 in RIME.

Barplots depicting the % peptide sequence coverage for ER, GATA3 and TET2 in separate RIME experiments, indicating that these three proteins are reciprocally detected as interactors of one another using RIME. Results represent mean ± SD (n=3). The number of unique peptides detected for each protein (the average of three biological replicates) is indicated above each bar.

TET2 has been identified previously as an ER interactor in RIME datasets from MCF7 cells (Papachristou et al. 2018), however this is the first time RIME for TET2 itself has been described. Again, given the detection of TET2 as an ER interactor, the robust overlap between the ER and TET2 interactomes provides some validation to the TET2 antibody used. Further confirming the capacity of the TET2 antibody to detect known TET2 interactors, published co-immunoprecipitation data has shown that TET2 interacts with p300 (Zhang et al. 2017), OGT (Chen et al. 2013; Deplus et al. 2013; Vella et al. 2013), and HDAC2 (Zhang et al. 2015), and all three of these proteins were co-purified with TET2 in RIME.

Beyond this, several additional interactors were detected in the RIME data that could shed light on the dynamic interplay between ER, GATA3 and TET2, and the potential contribution of TET2 to ER transcription. Amongst the other proteins detected as common interactors of ER, TET2 and GATA3 were the ER co-repressor proteins NCOR1 and NCOR2, and the scaffolding protein AKAP8. Common interactors of ER and TET2 included APEX1, an apurinic/apyrimidinic (AP) lyase thought contribute to active DNA demethylation through its role in base-excision repair (Hajkova et al. 2010), and several members of the MCM family of DNA helicases important for DNA replication during the cell cycle (MCM3 and 7). Collectively, these proteins may implicate TET2 in various processes including transcriptional control, scaffolding/mediation of protein-protein interactions, active regulation of DNA modifications (or DNA repair), and DNA replication during the cell cycle, all potentially as part of the ER complex.

Given the large overlap between the ER, GATA3 and TET2 interactomes, examination of proteins detected in each of these RIME experiments individually may also help to evaluate the potential co-functionality of these proteins in the ER complex. For example, the DNA methyltransferase DNMT1 was was co-purifed with both ER and GATA3, furthering the potential association between DNA methylation pathways and ER signalling. In relation to TET2, whilst other TET family members TET1 and TET3 are thought to be directed to GC-rich target sites via their CXXC domains, TET2 lacks this domain, hence it is unclear how it is targeted to specific genomic regions. One mechanism may be though interactions with other CXXC domain-containing proteins (Hu et al. 2013), and both CXXC4/IDAX and CXXC5/RINF have been shown to facilitate the association of TET2

with chromatin (Ko et al. 2013; Ma et al. 2017). CXXC5 was present as an ER interactor, suggesting that this protein might help mediate TET2-DNA interactions as part of its role in the ER complex. Finally, the E3 ubiquitin ligase UHRF2, shown to enhance interactions between TET proteins and hydroxymethylated DNA (Spruijt et al. 2013), was also detected in both ER and GATA3 RIME experiments, suggesting that TET2 may interact with 5hmC proximal to ER sites. These speculations may be tempered by the absence of CXXC5 and UHRF2 from the TET2 RIME, suggesting they might be part of a larger ER/GATA3 complex, but not necessarily via direct association with TET2 sub-complexes. However, the fact that these proteins are not detected as TET2 interactors could also be reflective of greater efficiency of the ER antibody compared to the TET2 antibody. Stoichiometry, positioning, and expression levels of certain bait proteins compared to others may also affect the subsets of interactors detected using different antibodies against proteins participating in the same complex, and the profile of interactors detected in the ER RIME compared to the TET2 RIME might also reflect these differences. Furthermore, RIME provides an aggregate view of all protein complexes immunoprecipitated from a heterogeneous population of cells, with no resolution of complex-to-complex variation. Whilst detection of the key interactors involved in the majority of complexes will likely be retained, resolution of interactions important only at subsets of these complexes may be weakened or lost, dependent on the above variables. This highlights how probing the same protein complex using several different bait proteins, as has been attempted here, may facilitate a more comprehensive picture of the proteins acting together in a cell.

3.2.4 TET2 antibody testing using ChIP-seq

Having confirmed that ER and TET2 interact as part of the ER complex, the chromatin profile of these interactions was next investigated using ChIP. Despite the similarities between the RIME and ChIP-seq protocols, anecdotal observations demonstrate that antibodies performing robustly in RIME do not always perform comparably in ChIP. Therefore, the TET2 antibodies previously tested using RIME, including several additional candidates, were screened in a single replicate of ChIP-seq in asynchronous, untreated MCF7 cells. ER ChIP-seq was performed as a control, and peaks were called using MACS (Zhang et al. 2008). The ER control ChIP-seq detected a robust number of binding events (34,692), demonstrating high signal and low background (Figure 3.8A, bottom track). Based on the RIME data indicating TET2 as a central member of the ER complex, and

the identical nature of the crosslinking conditions between the ChIP and RIME protocols, it was deemed plausible that TET2 binding may be detected at ER sites. This held true, and as such, snapshots of ChIP-seq tracks at two key ER-regulated genes, *RARA* and *GREB1* (Figure 3.8A) are provided as examples of TET2 antibody performance. Whilst Santa Cruz sc398535 was the poorest performer in terms of the number of binding events detected (with 975 peaks detected), Diagenode C15410255 (5,755 peaks), Abcam ab230358 (6,488 peaks), Bethyl A304-247A (5,521 peaks), and Pierce PA5-35847 (6,508 peaks) performed comparably. However, a far larger number of peaks (21,359) were detected using Abcam ab94580, more than three times those detected by the next-best performer, Pierce PA5-35847. Based on ChIP-seq signal, Abcam ab94580 was also the top-ranked antibody, displaying coverage approximately ten times that of Abcam ab230358, the next-best performer in this respect, at the sites shown in Figure 3.8A.

A Venn diagram indicating positional overlap of peaks for three of the best-performing antibodies based on a combination of their ChIP-seq performance and RIME protein sequence coverage (refer to Figure 3.4) is shown in Figure 3.8B. Whilst the binding sites detected using Pierce PA5-35847 formed a near-total subset of those detected by Abcam ab94580, the peaks detected using the Bethyl antibody A304-247A demonstrated a smaller overlap, with approximately ~50% of the peaks detected unique to this antibody. Although all TET2 antibodies demonstrated very similar binding profiles to ER, particularly at the *RARA* and *GREB1* enhancers depicted in Figure 3.8A, for Abcam ab94580 this similarity was striking, likely due to the improved coverage and signal-to-background ratio of this antibody. Although according to Figure 3.8B, the Bethyl, Pierce and Abcam antibodies all detected peaks that overlapped with ER binding, the peaks produced by both Abcam ab94580 and Pierce PA5-35847 formed a near-total subset of the total ER binding events, with 95% of the Pierce PA5-35847 peaks and 93% of the Abcam ab94580 peaks co-occupied by ER (corresponding 18% and 57% of ER peaks for each antibody, respectively). This high degree of ER and TET2 co-localisation is not unexpected given the interaction between ER and TET2 observed in the RIME data, and thus it was reasoned that the large number of ER-overlapping peaks detected by Abcam ab94580 may be valid and specific to TET2. For these reasons, Abcam ab94580 was selected for future ChIP-seq studies.

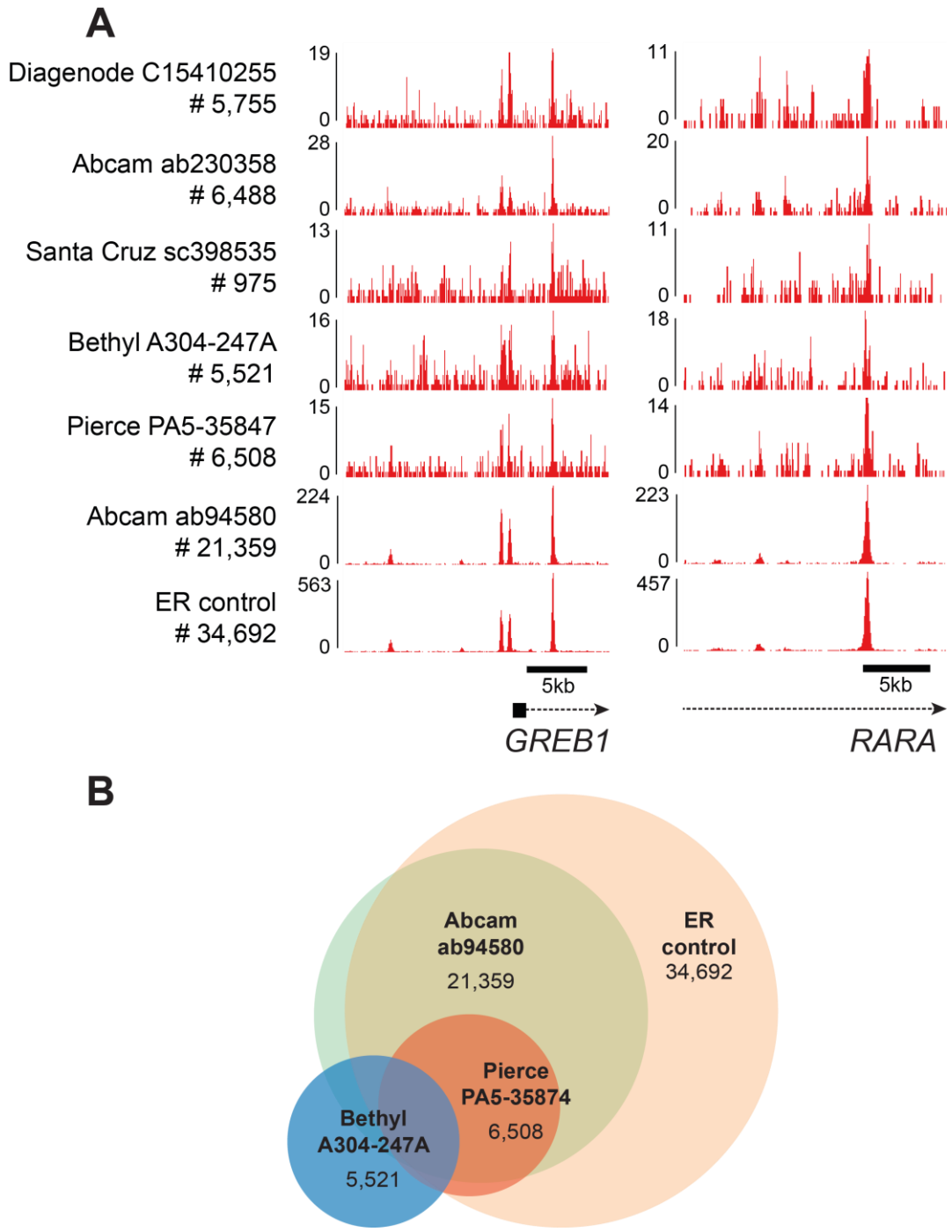


Figure 3.8. TET2 antibody testing using ChIP-seq.
A) UCSC genome browser tracks of ChIP-seq signal near two key ER target genes (*GREB1* and *RARA*) for the six TET2 antibodies tested, plus an ER control (performed using the ER antibody combination detailed in the Materials and Methods). All tracks are representative of one biological replicate, and all samples recieved equivalent (between 22 and 25 million) sequencing reads. The number of peaks called using MACS are indicated. B) Venn diagram showing positional overlap of binding sites for three of the best-performing antibodies based on a combination of ChIP-seq performance and RIME protein coverage (refer to Figure 3.3). Overlap with ER control peaks is also shown.

3.2.5 ER and TET2 co-localise at chromatin in ER+ cell lines and PDX models

The aim of the next investigations was to validate the preliminary findings described above, and to explore the interplay between ER and TET2 in several additional ER+ breast cancer models. ChIP-seq was first performed in two ER+ cell lines, MCF7 and ZR75-1. Asynchronous, untreated cells were used and four biological replicates were performed for each cell line. Peaks were called used MACS (Zhang et al. 2008) and only peaks occurring in all four independent replicates were considered in further analysis. This resulted in the identification of 16,884 and 13,423 TET2 binding sites in MCF7 and ZR71-1 cells, respectively (Figure 3.9). Comparison of TET2 and ER ChIP-seq in both cell lines confirmed the robust overlap of ER and TET2 binding observed in the preliminary experiments, with TET2 binding sites constituting a near-total subset of ER binding sites in both cell lines (Figure 3.9A). UCSC genome browser screenshots in Figure 3.9B demonstrate the overlap of ER and TET2 binding at regulatory sites for two key ER target genes, *RARA* and *GREB1*. Motif analysis of TET2 binding regions further confirmed the association between TET2 and ER, with ER and FOXA1 motifs the two most significantly enriched sequences within TET2 peaks (Figure 3.10).

The co-localisation of ER and TET2 was next validated in two ER+ patient-derived xenograft (PDX) models (Bruna et al. 2016). In these models, TET2 sites again formed a subset of ER binding sites (Figure 3.11A), with ER and TET2 co-bound in proximity to ER-regulated genes, as in the cell line models (Figure 3.11B). PDX model STG195 possesses an ER mutation (Y537S), suggested to promote constitutive ER activity and resistance to antihormone therapies (Harrod et al. 2017). The proportion of ER sites co-occupied by TET2 in this model was comparable to the proportion of ER sites bound by both factors in wild-type model AB555, implying that this mutation does not alter the interaction of TET2 with the ER complex. The high degree of genomic co-localisation of ER and TET2, and the near-absence of ER-independent TET2 binding sites is thus observed across all four independent ER+ breast cancer models.

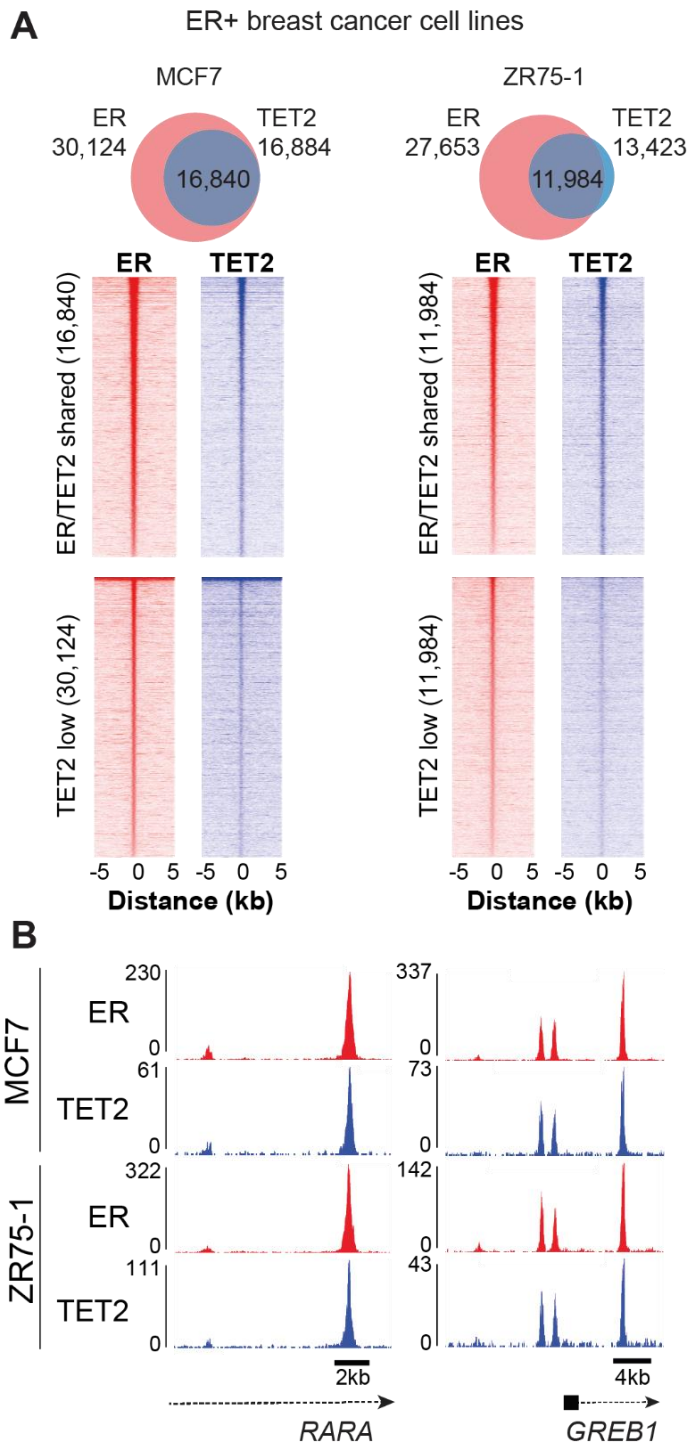


Figure 3.9. TET2 and ER interact at chromatin in ER+ breast cancer cell lines.
A) Venn diagrams indicate positional overlap of ER and TET2 ChIP-seq peaks in ER+ breast cancer cell lines MCF7 and ZR75-1. The heatmaps below each Venn diagram illustrate the ChIP-seq signal intensity for ER and TET2 at ER/TET2 shared sites (top), and "TET2 low" sites where TET2 peaks were not called (bottom). B) UCSC genome browser tracks demonstrate overlap of TET2 and ER peaks at ER target genes *RARA* and *GREB1*. For the schematics below each track, dotted lines indicate introns, boxes indicate exons, and arrowheads indicate direction of transcription. ChIPs were perfomed in quadruplicate.

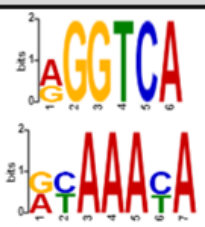
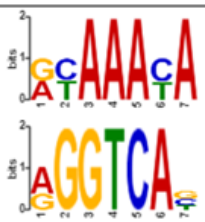
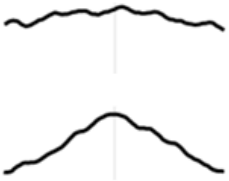
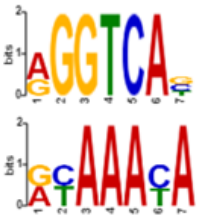
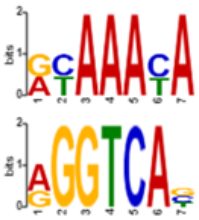
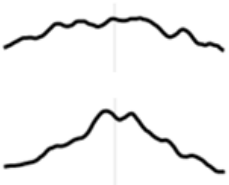
ChIP	Top 2 enriched motifs	E-value	Distribution	Similar motifs
MCF7 ER		$9.7e^{-1189}$ $3.3e^{-801}$	 Not Centrally Enriched	ESR1 FOXA1
MCF7 TET2		$3.6e^{-712}$ $3.4e^{-611}$		FOXA1 ESR1
ZR75-1 ER		$1.5e^{-867}$ $1.6e^{-801}$		ESR1 FOXA1
ZR75-1 TET2		$4.6e^{-575}$ $5.1e^{-364}$		FOXA1 ESR1

Figure 3.10. *De novo* motif analysis of ER and TET2 peak regions reveals enrichment of ESR1 and FOXA1 motifs.
Enrichment analysis was performed using MEME Suite tools (Machanick and Bailey 2011). The top two enriched motifs according to E-value (determined by significance of the motif according to the discovery program reporting the motif) are shown in column 2. Column 4 shows the distribution of the best matches to the enriched motif within the peak regions analysed (the vertical line corresponds to the centre of the sequences). Mean, median, and range (in bp) for inputted sequences for each ChIP-seq are (in order): MCF7 ER – 605, 527, 4160; MCF7 TET2 – 421, 505, 4631; ZR75-1 ER – 592, 505, 4631; ZR75-1 TET2 – 403, 352, 2688. Specific identities of motifs bearing similarity to those in column 2 are displayed in column 5, demonstrating that these motifs represent ER and FOXA1 binding regions.

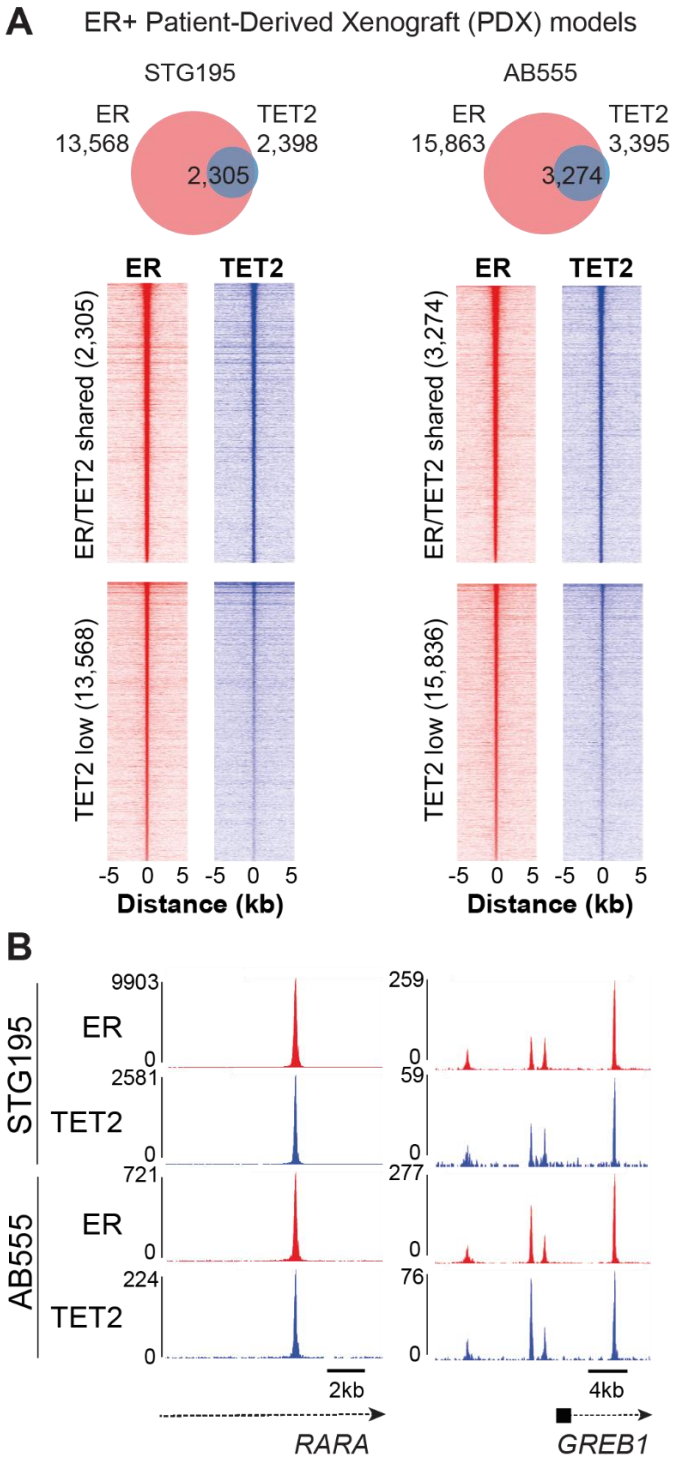


Figure 3.11. TET2 and ER interact at chromatin in ER+ PDX models.
A) Venn diagrams indicate positional overlap of ER and TET2 ChIP-seq peaks in two ER+ PDX models, STG195 and AB555. The heatmaps below each Venn diagram illustrate the ChIP-seq signal intensity for ER and TET2 at ER/TET2 shared sites (top), and "TET2 low" sites where TET2 peaks were not called (bottom). B) UCSC genome browser tracks demonstrate overlap of TET2 and ER peaks at ER target genes *RARA* and *GREB1*. For the schematics below each track, dotted lines indicate introns, boxes indicate exons, and arrowheads indicate direction of transcription. ChIPs were perfomed in quadruplicate.

When cross-comparing ChIP-seq between different factors, variations in antibody efficiency and specificity (discussed in the context of RIME in section 3.2.3) can limit definitive conclusions about true patterns of genomic overlap. Validation of the TET2 antibody used here against several other candidates in both ChIP and RIME suggests it is specific and effective. However, it is likely that TET2 interacts with chromatin at a wider range of sites than those detected in the above models, but that these are lower affinity sites and hence are not captured using ChIP. While it is important to bear in mind the caveats of antibody-based techniques, the lack of robust numbers of ER-independent TET2 sites could also suggest a role for ER in targeting TET2 to chromatin. As described previously, TET2 lacks the CXXC domain possessed by other TET family members, hence how it is directed to specific genomic regions is uncertain. Investigations in several model systems have proposed different proteins that may determine site-specificity of TET2-chromatin interactions (Chen et al. 2018; Hassan et al. 2017; Wang et al. 2015). From these studies, it seems likely that if TET2 does require intermediate proteins to facilitate its targeting to specific sites, these mediators are likely to vary with both cell, and potentially genomic, context. These ChIP-seq results, demonstrating close association of the TET2 chromatin binding profile with that of ER, build on the RIME findings and imply ER may help target TET2-chromatin interactions in an ER+ breast cancer context. This idea is explored further in Chapter 4.

In all four models (two cell lines and two PDX tumour samples), genomic annotation of ER sites (Figure 3.12) revealed binding of ER mainly to promoter-distal sites, as expected (Carroll et al. 2006). TET2 embodied this preference for distal sites even further, exhibiting a smaller fraction of promoter-proximal binding sites than ER in all four models. This finding that TET2 may preferentially associate with enhancers is consistent with recent literature (Rasmussen et al. 2015, 2019), a turning point from previous paradigms where, in the absence of high-quality TET2 ChIP-seq data, it was assumed TET2 occupancy would bias towards promoters, in line with observations for TET1 and TET3 (Jin et al. 2016; Williams et al. 2011; Wu & Zhang 2011). These endogenous TET2 mapping approaches show that in ER+ breast cancer cells, the binding of TET2 tracks that of the driving transcription factor, ER.

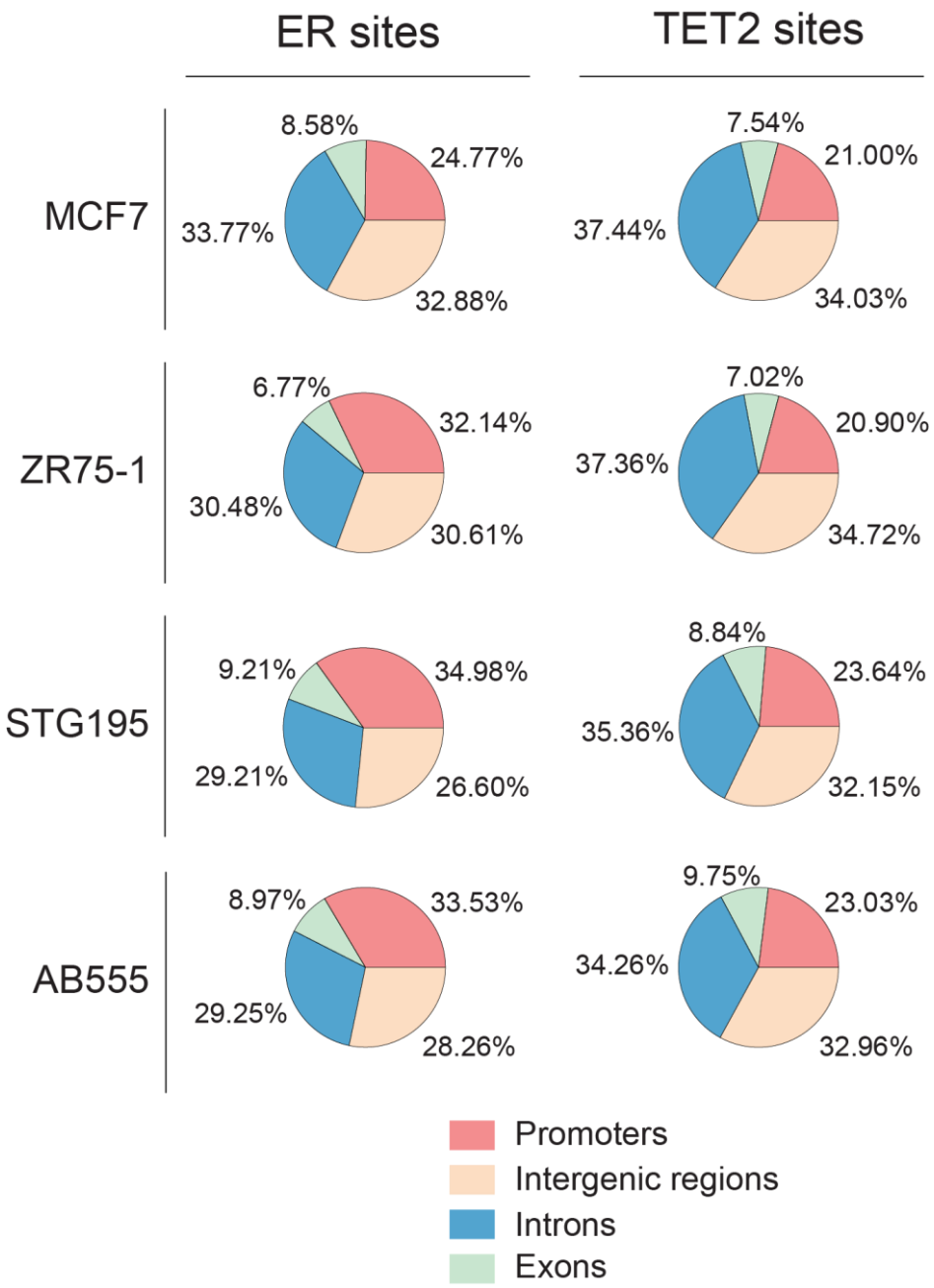


Figure 3.12. Genomic annotation of ER and TET2 sites indicates binding to a large fraction of promoter-distal regions.
Pie charts show classification of ER and TET2 binding sites according to genomic location for all the models tested. Promoters were defined as regions inclusive of 1 kb downstream and 2 kb upstream of the TSS.

In addition to confirming that ER and TET2 co-bind at key ER target genes, two examples of which (*GREB1* and *RARA)* are shown in Figures 3.9B and 3.11B, inspection of individual ChIP-seq tracks indicated that the TET2 gene itself possesses ER binding sites 20-30 kb upstream of its TSS in all four ER+ breast cancer models (Figure 3.13). This indicates that TET2 may be an ER target gene in ER+ breast cancer, and is supported by recent studies in MCF7 cells demonstrating that TET2 expression is induced by oestrogen (Wang et al. 2018) and robustly repressed by tamoxifen (Papachristou et al. 2018). In combination with the finding that TET2 expression is reduced by GATA3 knockdown, these results indicate that TET2 is a common target gene of both ER and GATA3 that is a central component of the ER complex on chromatin.

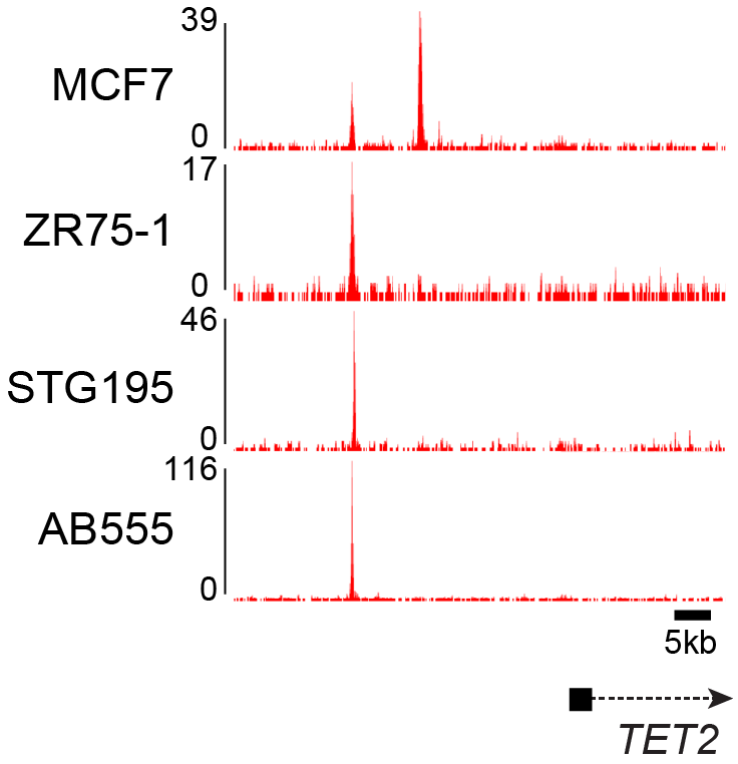


Figure 3.13. ChIP-seq shows ER binding sites distal to the TET2 promoter.
UCSC genome browser tracks demonstrating ER binding sites upstream of the TET2 TSS in two ER+ breast cancer cell lines (MCF7 and ZR75-1) and two ER+ PDX models (STG195 and AB555). ChIPs were performed in quadruplicate.

3.2.6 TET2 expression is associated with outcome in breast cancer

As described in Chapter 1, TET2 is rarely mutated in breast cancer (Scourzic et al. 2015; Stephens et al. 2012). Nevertheless, reduced TET2 expression is observed in breast tumours compared to matched normal tissue (Yang et al. 2013), and examination of a cohort of 162 breast cancer samples by Yang et al. (2015) revealed that lower levels of TET1, TET2, and TET3 mRNAs are associated with poor prognosis. However, neither of these studies discriminated breast cancers based on molecular subtype. To investigate whether TET2 associates with clinical outcome specifically in ER+ breast cancer, the online tool KM plotter (Győrffy et al. 2010) was used to assess the relationship between TET2 mRNA expression and relapse-free survival (RFS) in ER+ and ER- breast cancer patients (Figure 3.14). Higher expression of TET2 was significantly associated with improved relapse-free survival in the ER+ cohort, but not in the ER- cohort. Although further validation using additional datasets would be required to confirm this observation (for example to match the numbers of ER+ and ER- patients assessed), this demonstrates that TET2 predicts clinical outcome when ER+ disease is assessed exclusively.

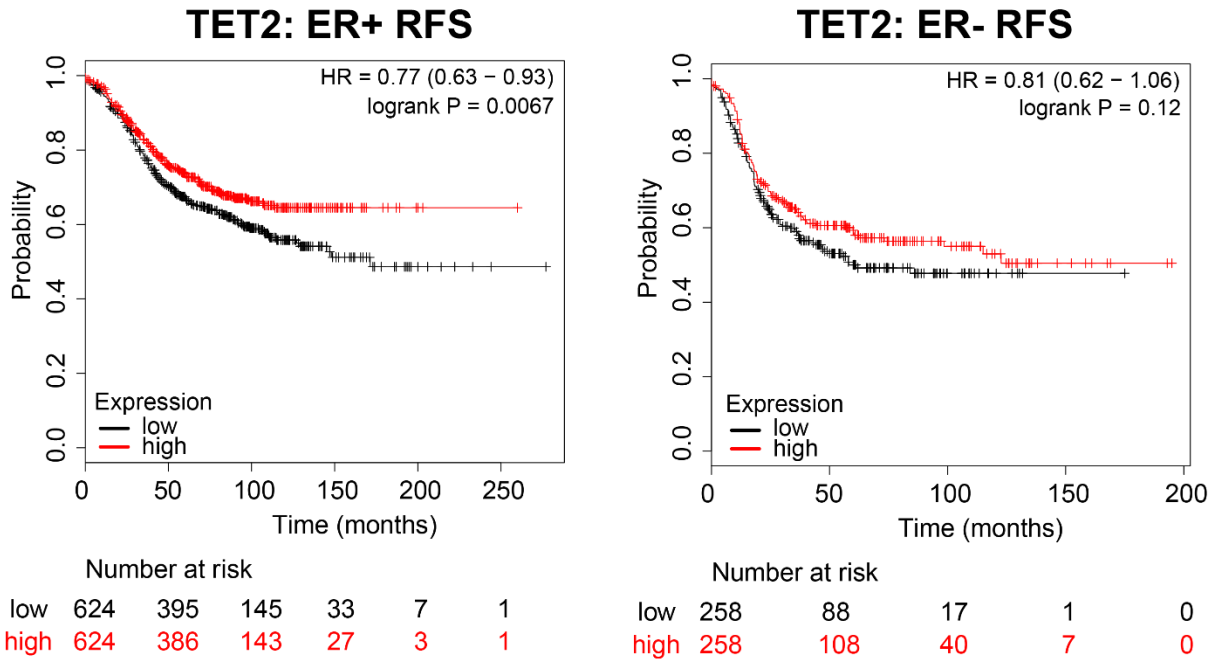


Figure 3.14. Higher TET2 mRNA expression is associated with improved relapse-free survivial in ER+ breast cancer.
Kaplan-Meier plotter (Győrffy et al. 2013) was used to assess relapse-free survival (RFS) in ER+ and ER- breast cancer patients. HR = hazard ratio.

3.3 Summary

Through the use of qPLEX-RIME, a recently developed proteomic tool for investigating endogenous interacting proteins, the findings in this chapter indicate that GATA3 regulates the composition of the ER complex, shedding new light on its role in ER biology. Amongst the proteins significantly enriched in the ER complex upon loss of GATA3 were the transcription factors LHX4 and ZBTB34, whilst the dioxygenase enzyme TET2 was significantly depleted. It has been suggested that LHX4 and ZBTB34 are capable of interacting with methylated DNA (5mC) (Filion et al. 2006; Qi et al. 2006; Yin et al. 2017), and the dioxygenase TET2 has a role in DNA methylation pathways through its function in oxidising 5mC to further DNA modifications, a process thought to be central to DNA demethylation. The common functional thread unifying these proteins thus suggests that GATA3 may have an indirect role in regulating levels of both DNA methylation and oxidised forms of this mark, and how these modifications are interpreted by the ER complex. The fact that TET2 was depleted from the ER complex in response to GATA3 knockdown presented it as a more tractable target to study in relation to breast cancer than the enriched factors LHX4, ZBTB34 or GREB1L, which are not detected as ER interactors under normal conditions in these cells. In addition, the presence of a wide body of existing research on TET2 and its functions, in particular the developing research on its role in transcription, prompted further investigation of the role of this factor in the ER complex. This work forms the basis of the remainder of this thesis.

Due to the lack of well-optimised, commercially available antibodies for immunoprecpiptation of TET2, both RIME and ChIP-seq were used to evaluate a range of antibody options. Upon selection of a candidate for robust detection of TET2, TET2 was confirmed as central to the ER complex through detection of its reciprocal interaction with both ER and GATA3 in RIME. This was further evidenced by the extensive overlap between TET2, GATA3 and ER interactors, including many factors salient to ER biology such as FOXA1, RARα and CARM1. ChIP-seq in several ER+ breast cancer models, representing both cell line models and PDX tissue, demonstrated that TET2 binds to a subset of ER regulatory sites, and additionally that TET2 itself appears to be an ER target gene, with a putative ER enhancer 20-30 kb upstream of its TSS. This reinforces the notion of a functional role for TET2 with regards to ER signalling, as it has been shown that the ER complex can regulate its own activity through controlling transcription of its

core components. For example, GREB1 and RARα, both proteins linked to ER function, are also ER-regulated genes (Mohammed et al. 2013; Liu et al. 2014; Ross-Innes et al. 2010). Furthermore, the strong link between ER and TET2 chromatin occupancy, and the observation of very few ER-independent TET2 binding sites, suggest a possible role for ER in targeting TET2 to chromatin. This idea is explored further in Chapter 4.

The suggestion of a functional link between ER and TET2 is a novel concept in the ER field. Whilst this work was in progress, Wang et al. (2018) established a link between ER and TET2, showing that these proteins co-localise on chromatin in MCF7 cells, and proposing that TET2 has a role in the expression of ER target genes. The work in this chapter builds on this ER-TET2 connection by demonstrating that the participation of TET2 in the ER complex is regulated by the key ER cofactor GATA3. This work also shows, using RIME, that ER, GATA3 and TET2 interact as part of a protein complex, and reinforces the connection between TET2 and ER through demonstrating their interaction at chromatin in several additional ER+ breast cancer models. In terms of the importance of TET2 in breast cancer, existing studies have linked increased TET2 expression with improved prognosis (Wang et al. 2018; Yang et al. 2015). The findings that TET2 forms a key part of the ER complex, and itself appears to be an ER target gene, suggest this improved prognosis may be partly due to a requirement of TET2 for the formation of a fully functional ER complex that is sensitive therapeutic targeting. Further in-depth studies will be required to validate this observation.

For the work in this chapter, RIME analysis was performed by the Proteomics Core Facility, with further bioinformatic analysis performed by Dr Kamal Kishore of the Bioinformatics Core Facility. Sequencing was performed by the Genomics Core Facility, and RNA-seq and ChIP-seq data processing and analysis was performed by Dr Igor Chernukhin.

Chapter 4

The role of TET2 in gene expression, and the interdependence of ER/TET2-chromatin interactions

4.1 Introduction

The work in Chapter 3 demonstrated TET2 as a key component of the ER complex, and showed that ER and TET2 co-localise at chromatin in a range of breast cancer models, implying that there may be a functional relationship between these two proteins. Breast cancers driven by ER are associated with improved clinical outcome compared to ER- subtypes, partly due to fact that the ER complex presents a tractable target for antihormone therapies (Parker et al. 2009). The hypothesis that TET2 might be important for the proper functioning of the ER complex is corroborated by the finding that lower levels of TET2 are associated with poor prognosis in breast cancer (Yang et al. 2015), similar to the association between ER expression and outcome.

Based on these findings, the potential relationship between ER and TET2 in transcriptional control was explored further. Firstly, the TET2-regulated gene programme was investigated in ER+ breast cancer cells to determine whether this links to aspects of the ER programme known to contribute to a malignant phenotype. Reports in the literature combined with the work in Chapter 3 show that TET2 is an ER-regulated gene (Papachristou et al. 2018; Wang et al. 2018), and several oestrogen-regulated members of the ER complex have been shown to be important for ER-chromatin interactions and expression of ER target genes, including FOXA1, GREB1 and NCOA3 (AIB1) (Hurtado et al. 2011; Mohammed et al. 2013; Tikkanen et al. 2000). Based on this, and given the observation that ER and TET2 co-localise at chromatin at a large fraction of ER sites as

observed in Chapter 3, the requirement of TET2 for ER-chromatin interactions was also examined.

As mentioned in Chapter 1, TET1 and TET3 possess CXXC domains, structural features known to direct targeting of proteins to various CG-rich sequences (Xu et al. 2018). Uniquely to the TET family, TET2 lacks this domain, hence how it is targeted to specific regions of the genome remains a topic of research. It has been suggested that other CXXC domain-containing proteins may instead co-operate with TET2 to direct its chromatin interactions. One such protein, CXXC5, shown to facilitate recruitment of TET2 to promoters in plasmacytoid dendritic cells (pDCs) (Ko et al. 2013), was detected as a component of the ER complex in Chapter 3, implying a potential association between TET2 and CXXC5 with respect to ER signalling. Nevertheless, TET2 appears to preferentially target enhancers, in contrast to TETs 1 and 3 which favour promoters (Jin et al. 2016; Rasmussen et al. 2015, 2019; Williams et al. 2011). This distinction, coupled with the particular importance of enhancers in cell type-specific functions, further reinforces that TET2 may rather employ context-dependent, and potentially CXXC-independent mechanisms to access its target sites. Several studies have revealed insights into how TET2 might be recruited to chromatin, with Wang et al. (2015) showing that the transcription factor WT1 recruits TET2 to regulate gene expression in the human leukaemia cell line HL-60. Furthermore, Chen et al. (2018) showed a role for SNIP1 in bridging the interaction between TET2 and the transcription factor c-MYC, aiding expression of c-MYC target genes in U2OS cells. This implies that TET2-chormatin interactions may be co-ordinated by several intermediary proteins at once. The extent of the TET2 and ER genomic co-localisation observed in Chapter 3 implies that ER itself might have a role in targeting of TET2 to chromatin in the ER+ breast cancer context. A potential role for ER in directing and stabilising TET2-chromatin interactions was therefore also explored in this chapter.

4.2 Results

4.2.1 siRNA-mediated TET2 knockdown in MCF7 cells

To examine the role of TET2 in ER+ breast cancer cells, its expression was silenced using siRNA. Asychronous MCF7 cells were transfected with siRNA targeting TET2, or a non-targeting siRNA control, and TET2 knockdown was confirmed as effective and stable at the mRNA level over a time course of 96 hours after a single transfection (Figure 4.1).

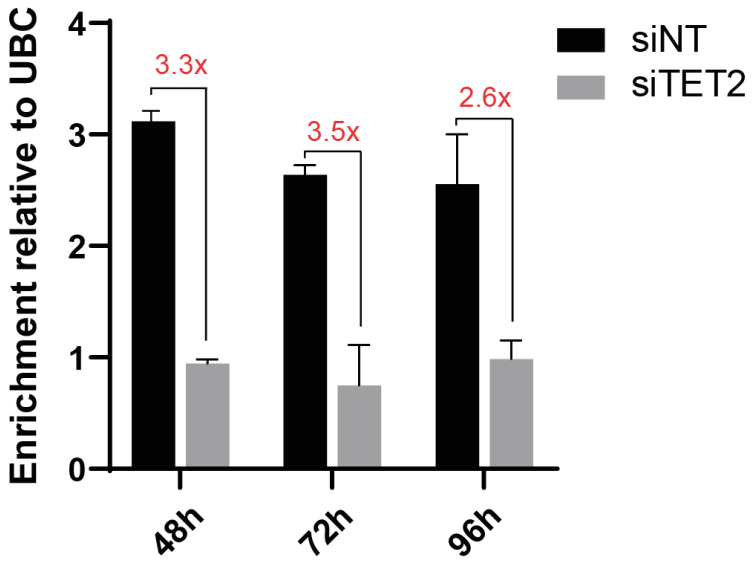


Figure 4.1. Confirmation of depletion of TET2 mRNA in MCF7 cells.
MCF7 cells were treated with an siRNA pool (10 nM) targeting TET2 (siTET2), or a non-targeting control (siNT). Cells were harvested at 48, 72 and 96 hours after transfection and TET2 mRNA levels were assessed using qRT-PCR. Results represent mean ± SD of enrichment relative to a housekeeping control gene (UBC) (n=2). * = $p \leq 0.05$, ** = $p \leq 0.01$.

4.2.2 Assessment of TET2 protein levels after TET2 knockdown

To confirm effective TET2 knockdown at the protein level, antibodies were tested for TET2 Western blot. Although several publications demonstrate Western blotting for TET2, a range of different antibodies have been used across the literature, probing TET2 in various different cell lines, so several candidates were assessed in order to find the best approach

for detecting TET2 in MCF7 cells. In total, seven antibodies were tested (see Materials and Methods), and multiple transfer methods applied, including dry and wet transfers of varying durations. However, no TET2-specific band could be detected in either whole-cell or nuclear-enriched lysates based on the comparison between control and TET2 knockdown samples. Figure 4.2A shows selected example Western blots.

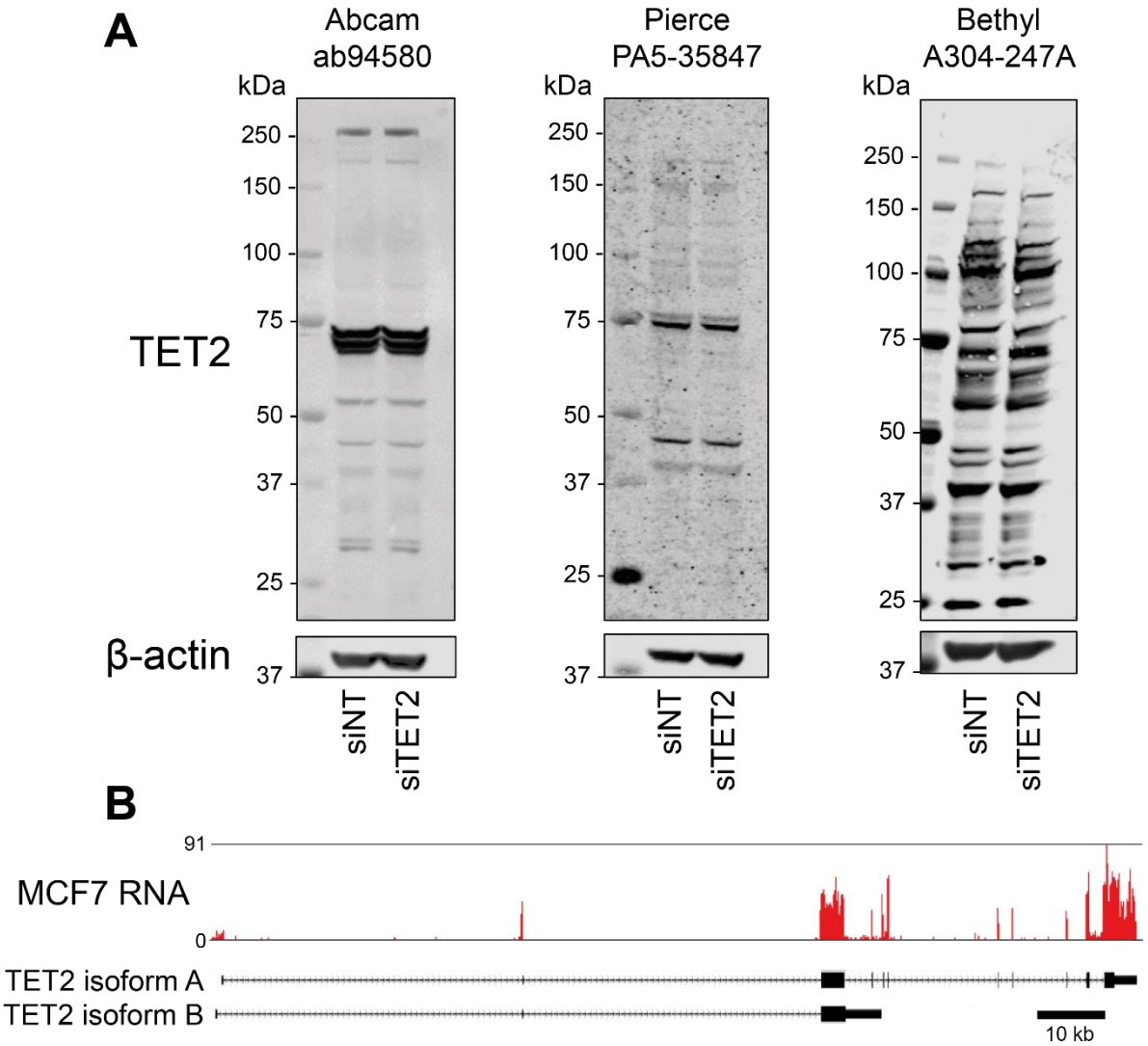


Figure 4.2. TET2 isoform expression in MCF7 cells and example TET2 Western blots.
A) Three example Western blots demonstrating the lack of specificity of several of the TET2 antibodies tested using MCF7 whole cell lysates. The blots shown were produced using the dry transfer system iBlot2 (see Materials and Methods). 50µg protein was loaded per blot. β-actin was used as a loading control. B) UCSC genome track of an RNA-seq library showing expression of TET2 in MCF7 cells. RefSeq schematics for TET2 isoform A (NM_001127208.2; ~223 kDa) and isoform B (NM_017628.4; ~130 kDa) are shown. Lines indicate introns, boxes indicate exons, and arrowheads indicate direction of transcription.

According to the RefSeq database (O'Leary et al. 2016), both the canonical long isoform of TET2 (isoform A, ~223 kDa) and a short isoform proposed to lack the TET catalytic domain (isoform B, ~130 kDa) (Lou et al. 2019) are expressed from the TET2 locus. According to the ER qPLEX-RIME data from Chapter 3, TET2 peptides depleted from the ER complex in response to GATA3 loss in MCF7 cells included those exclusive to the longer isoform of TET2, indicating this as the primary isoform of interest in these investigations (data not shown), and RNA-seq data further confirmed expression of this longer isoform at the RNA level in MCF7 cells (Figure 4.2B). As this longer isoform of TET2 has a high molecular weight (223 kDa), it is possible that this property might impair its transfer from gel to membrane. However, the use of multiple blotting approaches suggest that low expression of TET2 in MCF7 cells combined with the lack of specificity of these antibodies in Western blot are likely to have limited the effective detection of this protein. Given that TET2 was successfully immunoprecipitated for RIME and ChIP-seq using Abcam antibody ab94580, an alternative approach could have been to enrich for TET2 using this method prior to Western blotting. However, it was reasoned that improved sensitivity and accuracy in detection of TET2 might be achieved through collaboration with the Proteomics Core Facility at CRUK-CI. Both full proteome and protein-targeted mass spectrometry measurements offer improved sensitivity and accuracy over Western blotting, particularly for proteins exhibiting low expression (Liebler & Zimmerman 2013). Therefore, a combination of these techniques was applied to assess the protein-level effects of TET2 knockdown. TMT full proteome analysis, similar to qPLEX-RIME, relies on TMT labelling allowing quantitative comparisons between experimental conditions, but in an antibody-independent manner. Using this method, robust depletion of TET2 protein levels was confirmed in response to 72 hours of TET2 siRNA treatment, with no effect measured on total ER or GATA3 levels (Figure 4.3). For all subsequent assessments of TET2 protein levels in this work, Parallel Reaction Monitoring (PRM) was employed. This targeted proteomics approach allows quantification of proteins based on the measurement of specific peptides, unique to the protein of interest, using tandem mass spectrometry. Peptides are fragmented to generate a characteristic ion fragmentation pattern that indicates the targeted peptide sequence with high specificity. The signal intensity of these resulting fragments is compared to that of stable isotope-labelled reference peptide standards, and this serves as the basis for quantitative comparison of protein levels across different biological samples.

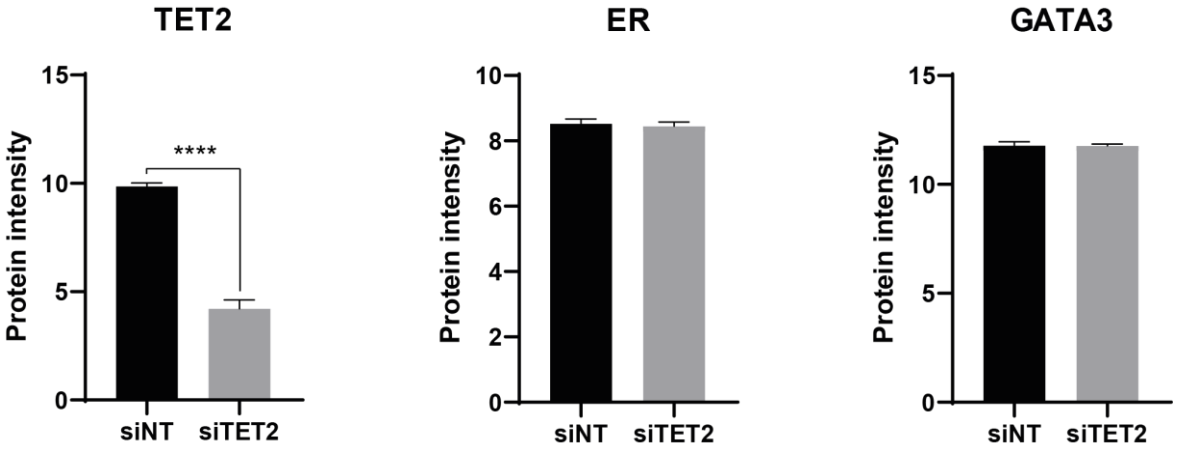


Figure 4.3. TET2 knockdown significantly depletes TET2 protein levels in MCF7 cells, with no effect on total ER or GATA3 protein levels.
MCF7 cells were treated with an siRNA pool (10 nM) targeting TET2 (siTET2) or a non-targeting control (siNT). Cells were harvested at 72 hours after transfection and full proteome analysis performed. Four replicates of each condition were included in an 11plex TMT MS run. Results represent mean ± SD protein intensity, the aggregate of intensities of the individual unique peptides identified for each protein. **** = $p \leq 0.0001$.

4.2.3 TET2 depletion affects proliferative pathways in MCF7 cells

To identify TET2-regulated genes, RNA-seq was performed following 48 hours of TET2 knockdown in asynchronous MCF7 cells. This revealed repression of 2,269 genes and activation of 2,144 genes ($p \leq 0.05$) (Figure 4.4). GATA3 mRNA expression was unaffected by TET2 silencing. ER mRNA expression was modestly repressed, although proteomic analysis after 72 hours of TET2 knockdown (Figure 4.3) suggested that this does not translate into an effect on total protein levels. To compare the gene regulatory programme of TET2 with those of ER and GATA3, RNA-seq was also performed 48 hours after knockdown of either ER or GATA3 in MCF7 cells. ER and GATA3 mRNA levels were robustly depleted, with ER transcripts reduced to 11% of control levels and GATA3 transcripts reduced to 20% of control levels by their respective silencing. As expected, TET2 mRNA levels were robustly and significantly repressed by both ER and GATA3 knockdown (to 23% of control levels by ER knockdown, and 43% of control levels by GATA3 knockdown). Overall, 60% of the genes significantly regulated by TET2 knockdown (2,656 out of 4,413 genes, $p \leq 0.05$) were also significantly regulated by both GATA3 and ER silencing. Given that TET2 appears to behave as a target gene of both

ER and GATA3, some of the concordance between the ER, GATA3 and TET2 gene regulatory profiles is therefore likely to be due to secondary effects as a result of TET2 depletion after ER and GATA3 silencing. Taking the 500 most induced and 500 most repressed genes in response to TET2 silencing (in terms of log2 fold change), 60% of these most differentially regulated genes were modulated in the same direction following ER knockdown (visually depicted in the heatmaps in Figure 4.5A). This included repression of key ER target genes such as *PGR*, *CCND1*, *XBP1*, and *CXCL12*. Importantly, ER/TET2 shared binding sites were enriched adjacent to TET2 target genes (Figure 4.6), further implying genomic co-operation between these two proteins. Interestingly, similar enrichment of ER/TET2 shared sites was observed both at genes induced by TET2 knockdown and those repressed by TET2 knockdown, implying that TET2 may be capable of both activating and repressive transcriptional activity as part of the ER complex. However, more detailed functional analysis would be required to confirm this observation.

To assess whether the genes regulated by TET2 show a tendency towards modulation of any particular pathway, gene ontology (GO) analysis was performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) (Huang et al. 2007; Huang et al. 2009). Genes repressed by TET2 knockdown showed significant enrichment of six functional categories linked to cell division and cell cycle processes (Figure 4.5B). In contrast, the genes induced by TET2 knockdown demonstrated significant enrichment of only two functional categories, linked to cell communication and signal transduction. When examining solely the TET2-regulated genes that also changed in response to GATA3 and ER knockdown, enrichment of the same functional categories was observed (Figure 4.5C). This demonstrates that the pathways most strongly affected by loss of TET2 in these cells are those converged on by GATA3 and ER regulatory processes, and are related to the cell cycle and proliferation. To place these findings into a functional setting, proliferation of MCF7 cells was assessed over a 4-day period after TET2 knockdown. Consistent with repression of cell cycle gene programs, MCF7 growth was inhibited by TET2 knockdown (Figure 4.7).

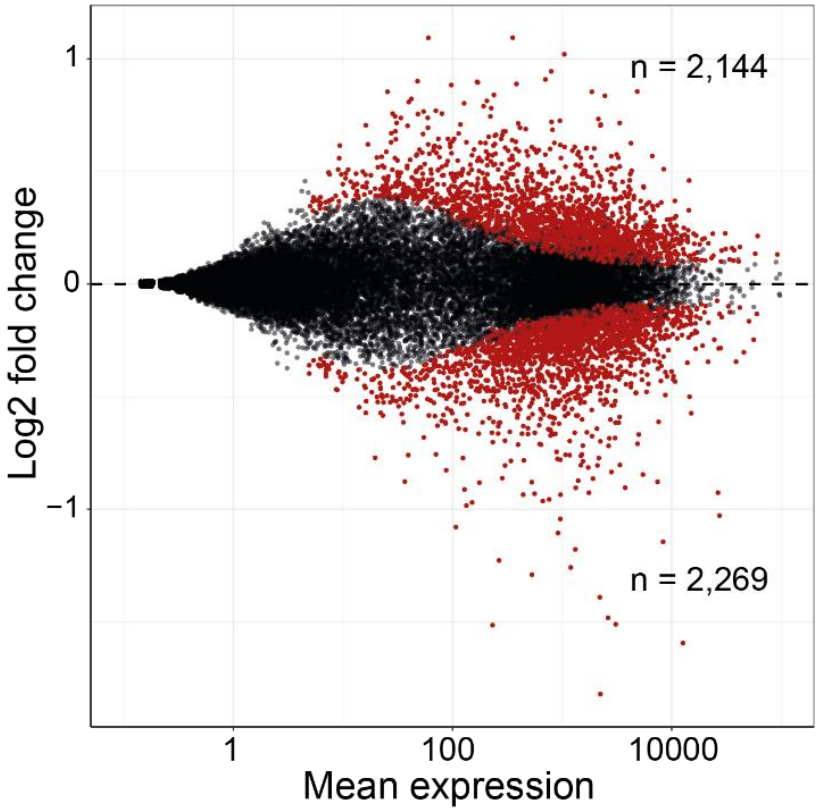


Figure 4.4. TET2 knockdown affects gene expression in MCF7 cells.
siRNA-mediated TET2 silencing (48 hours) resulted in repression of 2,269 genes and activation of 2,144 genes compared to control (n=6, $p \leq 0.05$, significantly regulated genes highlighted in red).

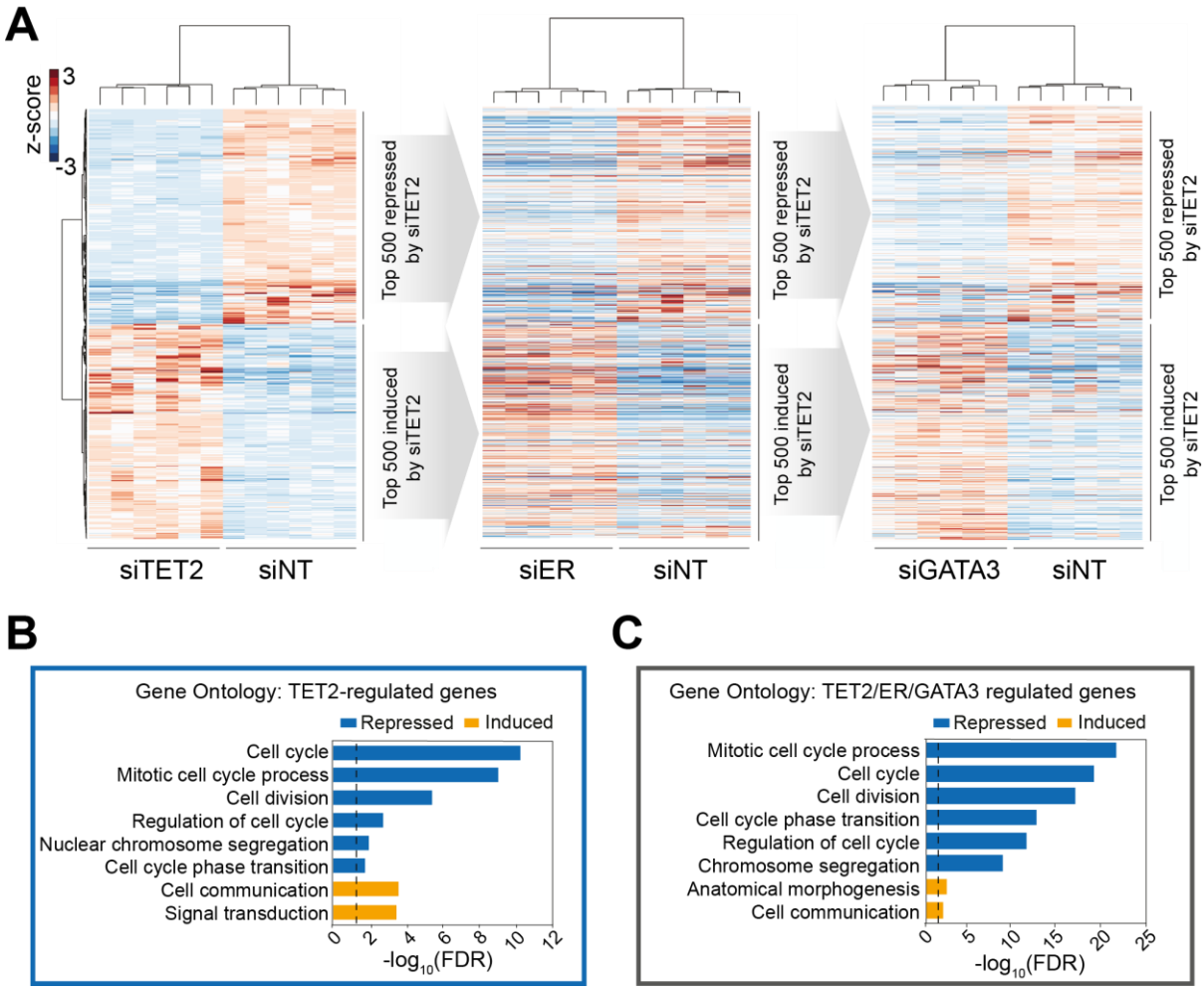


Figure 4.5. TET2 depletion affects gene pathways related to the cell cycle.
A) Heatmaps depict the top 500 induced and top 500 repressed TET2-regulated genes according to log2 fold change. Colour scale represents the relative expression (z-score) of genes across the two conditions (control and knockdown), calculated separately within each comparison (siTET2 vs siNT, siESR1 vs siNT and siGATA3 vs siNT). Hierarchical clustering of genes in the leftmost (siTET2) heatmap is preserved across all three heatmaps. Columns represent independent biological replicates (n=6). B) Barplot displaying –log10(FDR) for GO analysis of the top 500 induced and top 500 repressed TET2-regulated genes according to log2 fold change. Only categories with FDR ≤ 0.05 (threshold indicated by dotted line) are shown. C) Barplot displaying –log10(FDR) for GO analysis of the top 500 induced and top 500 repressed TET2-regulated genes according to log2 fold change, sub-selected from genes also significantly ($p \leq 0.05$) regulated by both GATA3 and ER silencing. The top 6 enriched categories are shown for repressed genes, and the top 2 for induced genes. Dotted line indicates FDR 0.05. Enriched processes were identified using the Biological Process category level 3 of the GO hierarchy (GOTERM_BP_3).

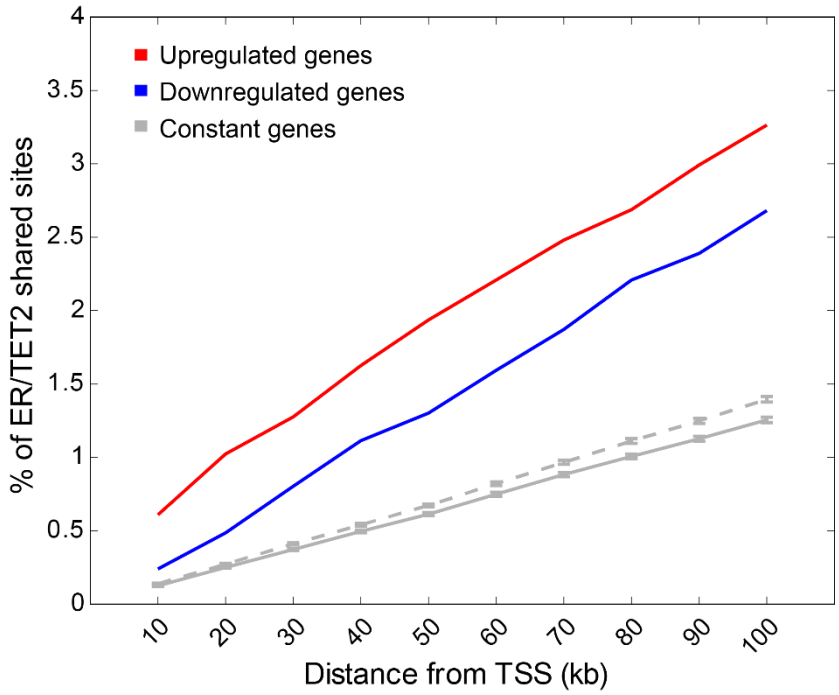


Figure 4.6. ER/TET2 shared sites are enriched in the vicinity of TET2-regulated genes compared to unchanging genes.
Graph shows the cumulative fraction of total ER/TET2 shared binding sites (n = 15,442, MCF7 cells) within up to 100 kb of the TSSs of three groups of genes: genes upregulated by TET2 knockdown (n = 2,144, red line), genes downregulated by TET2 knockdown (n = 2,269, blue line) ($p \leq 0.05$), and genes unchanging in response to TET2 knockdown (constant genes, grey lines). Constant genes were randomly selected from those with $p > 0.5$ and mean expression > 1.0. Grey lines indicate analysis based on constant genes: the dotted line indicates analysis matched to the number of downregulated genes, and the solid line indicates analysis matched to number of upregulated genes.

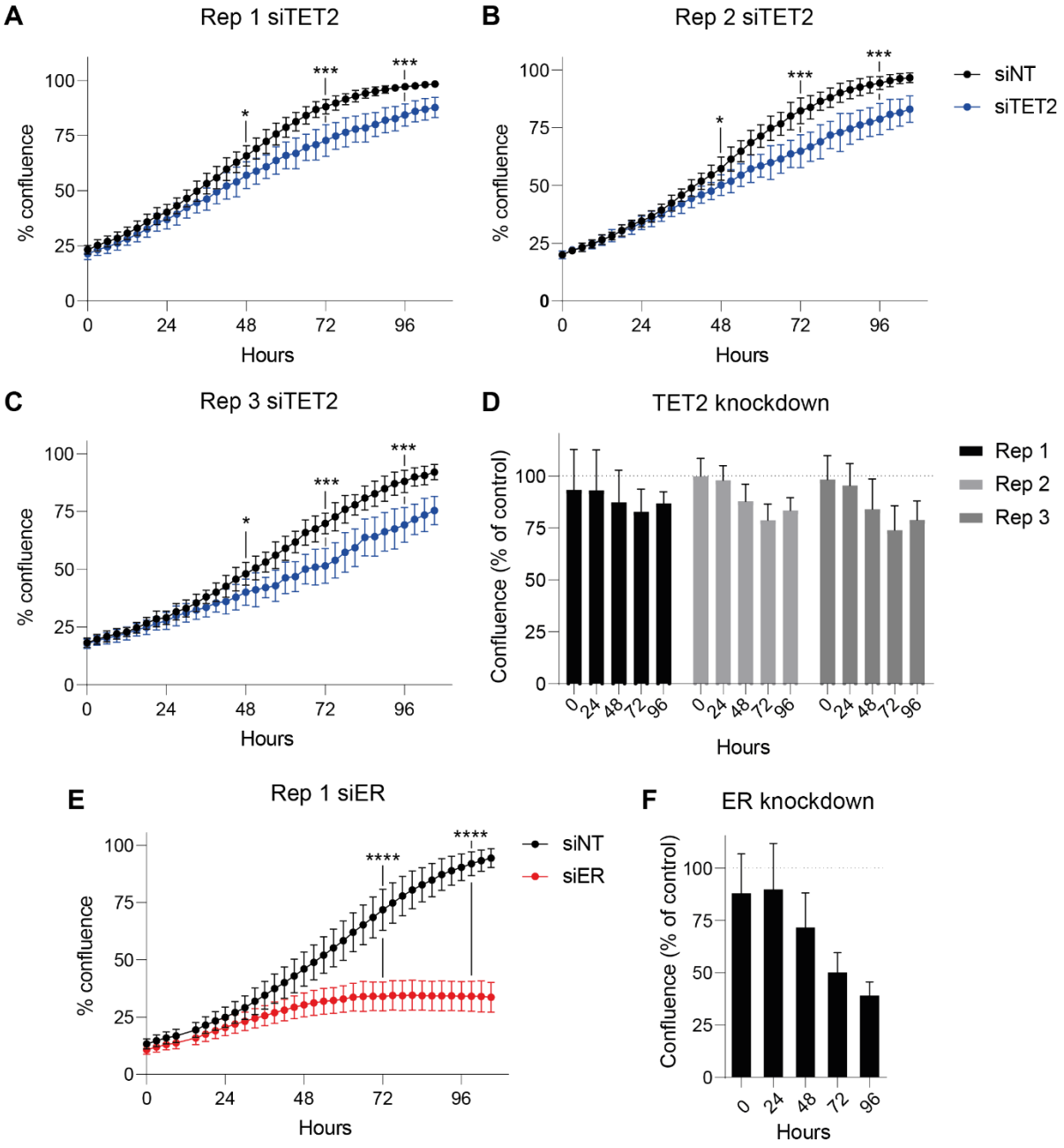


Figure 4.7. TET2 knockdown inhibits MCF7 cell growth.
Cell growth is expressed as mean ± SD of % confluence, assessed using the Incucyte Zoom™ system. Knockdown was performed at t = 0; cells were seeded 24 hours prior to this. Data points represent mean ± SD of ≥ 6 technical replicates (wells). Results for each biological replicate are plotted separately. ** = $p \leq 0.01$, *** = $p \leq 0.001$. siTET2 = TET2 knockdown, siER = ER knockdown, siNT = non-targeting control-treated cells. A), B) and C) show growth experiment replicates 1 to 3 demonstrating the effect of TET2 knockdown on growth of MCF7 cells. D) Barplot showing confluence of TET2 knockdown cells as a % of the confluence of corresponding control cells at 24, 48, 72 and 96 hours after treatment. E) Growth experiment demonstrating the effect of ER knockdown on MCF7 growth. F) Barplot showing confluence of ER knockdown cells as a % of the confluence of corresponding control cells at 24, 48, 72 and 96 hours after treatment.

4.2.4 Genes repressed by loss of TET2 are also repressed by tamoxifen treatment

As described in Chapter 3, higher expression of TET2 is associated with improved outcome in ER+ breast cancer. The finding that loss of TET2 inhibits the growth of ER+ MCF7 cells reinforces the idea that TET2 contributes to the proper function of the driving transcription factor, ER, which stimulates growth in these cells. The presence of a functional and hormone-sensitive ER complex is crucial to the success of targeted therapies for ER+ breast cancer. One of the current standard-of-care therapies, tamoxifen, binds ER and antagonises the effects of oestrogen on target genes (Clarke et al. 2001). To assess whether TET2 modulates growth through regulating similar proliferative pathways to those affected in the clinical setting, genes downregulated by TET2 depletion were compared with those repressed by tamoxifen treatment using data from Papachristou et al. (2018). As an ER target gene, TET2 is itself repressed by tamoxifen, with significant loss of both mRNA and protein observed after 24 hours' tamoxifen treatment according to data from Papachristou et al. (2018). However, acute tamoxifen treatment (6 hours) does not significantly affect TET2 protein levels, allowing the identification of gene expression changes in response to tamoxifen treatment only. Within genes significantly repressed by TET2 knockdown ($p < 0.05$), a subset of 175 genes was identified that are also acutely and significantly repressed by tamoxifen (6 hours 100 nM tamoxifen, $p < 0.01$). GO analysis using DAVID revealed that this subset of genes was enriched for functions relating to the 'cellular stress response' and 'DNA damage' pathways (FDR = $1.2e^{-7}$ and $2.4e^{-7}$, respectively) and several categories linked to cell cycle progression ('cell cycle phase', 'cell cycle process' and 'mitotic cell cycle', FDR = $1.4e^{-3}$, $2.3e^{-2}$ and $3.4e^{-3}$, respectively). These results further reinforce that TET2 regulates key ER target genes, including those targeted by agents proven to be effective in the clinic. In terms of therapy, an extension of this logic might be that targeting TET2 in ER+ breast cancer could provide clinical benefit in conjunction with current hormone therapies by potentiating their effects. However, it is still unclear precisely which aspects of TET2 function might be important for the ER complex, and furthermore, there are currently no known specific inhibitors of TET2 catalytic activity. Nonetheless, these studies illustrate that investigating the involvement of TET2 in ER function holds potential for developing a better knowledge of both basic and translational aspects of breast cancer biology.

Having established that TET2 is linked to expression of ER target genes, the next step will be to unravel the mechanisms behind this. As described in Chapter 1, TET2 oxidises methylated cytosines (5mC) to further DNA modifications 5hmC, 5fC and 5caC, with this process an intermediate step in DNA demethylation pathways (He et al. 2011; Inoue and Zhang 2011; Ito et al. 2011; Tahiliani et al. 2009). Dynamic regulation of these marks has been implicated in transcriptional control. If TET2 has a role in maintenance and turnover of these modifications at ER sites, this could contribute to the gene expression changes observed in response to loss of TET2. Whilst the work in this thesis was in progress, Wang et al. (2018) demonstrated that TET2 knockout MCF7 cells show changes in oestrogen-induced gene expression compared to wild-type cells. This was concurrent with differences in levels of methylation at ER binding sites in knockouts compared to controls, and this indeed may contribute to the changes observed in the acute knockdown system studied here. In addition, the TET2 RIME data discussed in Chapter 3, in conjunction with reports in the literature, suggest TET2 can interact with transcriptional regulators such HDAC2 and p300, as well as members of both the NCOA and NCOR families (Zhang et al. 2015; Zhang et al. 2017). This reinforces the idea that TET2 could have a role in stabilising transcriptional complexes in addition to its catalytic role regulating DNA modifications. The concept of TET2 providing stability to ER-chromatin interactions is explored in the sections below, whilst investigations into the role of TET2 in regulating DNA modifications at ER sites are carried out in Chapter 5.

4.2.5 TET2 knockdown affects global ER binding

To further investigate the relationship between TET2 and ER gene expression, the impact of TET2 loss on ER binding to chromatin was assessed. Firstly, TET2 ChIP-seq was performed after 48 hours of TET2 knockdown in asynchronous MCF7 cells to confirm that TET2 silencing effectively depletes TET2-chromatin interactions. Four biological replicates were performed, and peaks were called using MACS2 (Zhang et al. 2008). Valid peaks were defined as those occurring in all four replicates. A robust global reduction in TET2 binding was observed in siTET2-treated cells compared to controls (Figure 4.8). Specifically, 12,728 valid TET2 peaks were detected under control conditions, but only 3,098 TET2 peaks were called after TET2 silencing, validating the efficacy of the knockdown, and the specificity of the antibody.

To assess the effect of these depleted TET2-chromatin interactions on genome-wide ER binding, samples for ER ChIP-seq were harvested at 72 hours after TET2 siRNA treatment. A drop in overall ER binding was observed as a result of TET2 silencing (Figure 4.8). As demonstrated by the full proteome analysis in Figure 4.3, TET2 knockdown does not affect total ER protein levels. This result therefore strongly implies a role for TET2 in stabilising ER-chromatin interactions. Notably, despite the fact that TET2 binding was previously shown to occur at a large subset of, but not all, ER sites (Chapter 3), the profile of the MA plot in Figure 4.9A suggests that ER binding appears to be depleted, if modestly, at a large proportion of ER sites. This could suggest that the TET2 ChIP has not captured the full range of TET2 sites, and that TET2 is in fact present at many of the apparently TET2-independent ER sites, where it stabilises ER binding. Alternatively, this could imply that TET2 exerts secondary effects that stabilise ER-chromatin interactions in an indirect manner. Overall, these results imply that the loss of proper ER binding as a result of TET2 knockdown contributes to impaired regulation of ER target genes.

There are multiple mechanisms through which TET2 might impact ER-chromatin interactions. As previously mentioned, TET2 is a large protein (~223 kDa), and it is possible that it confers stability to the ER complex in a scaffolding capacity, recruiting additional cofactors and bridging connections between different parts of the complex. The absence of TET2 may therefore have a destabilising effect on the complex, contributing to the observed depletion of ER-chromatin interactions. A similar role has been suggested for ER cofactors RARα (Ross-Innes et al. 2010) and GREB1 (Mohammed et al. 2013), and the interaction of TET2 with a large number of other ER complex proteins and transcriptional regulators in RIME reinforces this idea. Interestingly, it has been shown that another TET family member, TET3, can stabilise the chromatin interactions of thyroid hormone receptors TRα1 and TRβ1, both of which are members of the nuclear receptor superfamily. Specifically, enrichment of both TRs in the chromatin fraction of HEK293T cells was dependent on the expression of TET3 (Guan et al. 2017). Notably, whilst a truncated form of TET3 lacking the catalytic and CXXC domains failed to achieve the same effect, the stabilising effect on TR-chromatin interactions was maintained in the presence of a TET3 mutant lacking dioxygenase ability, suggesting the stabilising function of TET3 may be independent from its catalytic role.

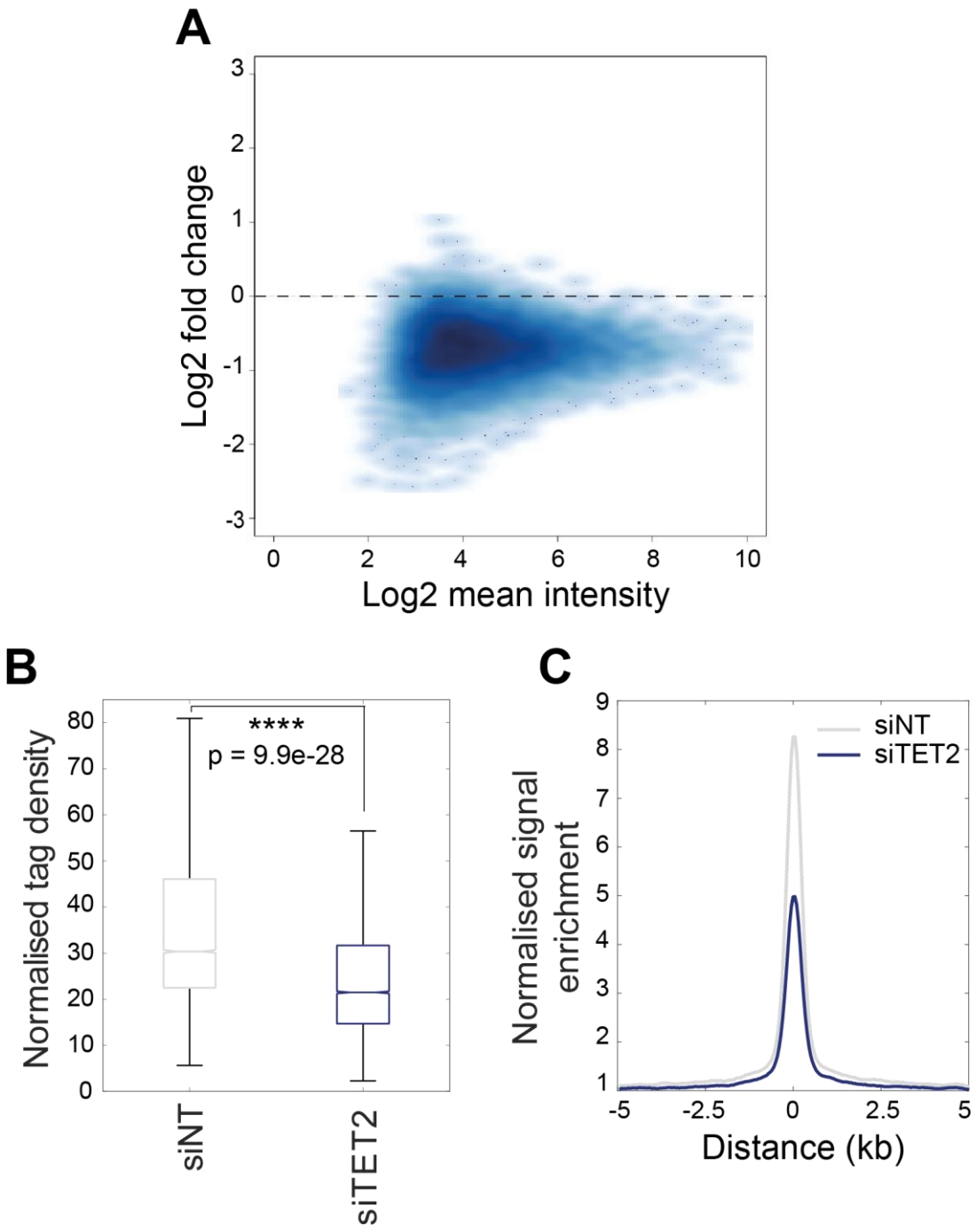


Figure 4.8. TET2 knockdown depletes global TET2 binding.
A) MA plot showing log2 fold change in TET2 binding in control versus TET2 knockdown conditions against log2 mean intensity of ChIP-seq signal for all TET2 sites (12,728 peaks). B) Boxplot showing the normalised tag density of TET2 ChIP-seq signal in control (non-targeting control siRNA, siNT) and TET2 knockdown (siTET2) conditions within all TET2 peaks. **** = $p \leq 0.0001$. C) Average plot showing normalised signal enrichment of TET2 ChIP-seq under control (siNT) or TET2 knockdown (siTET2) conditions within all TET2 peaks. TET2 knockdown was performed for 48 hours.

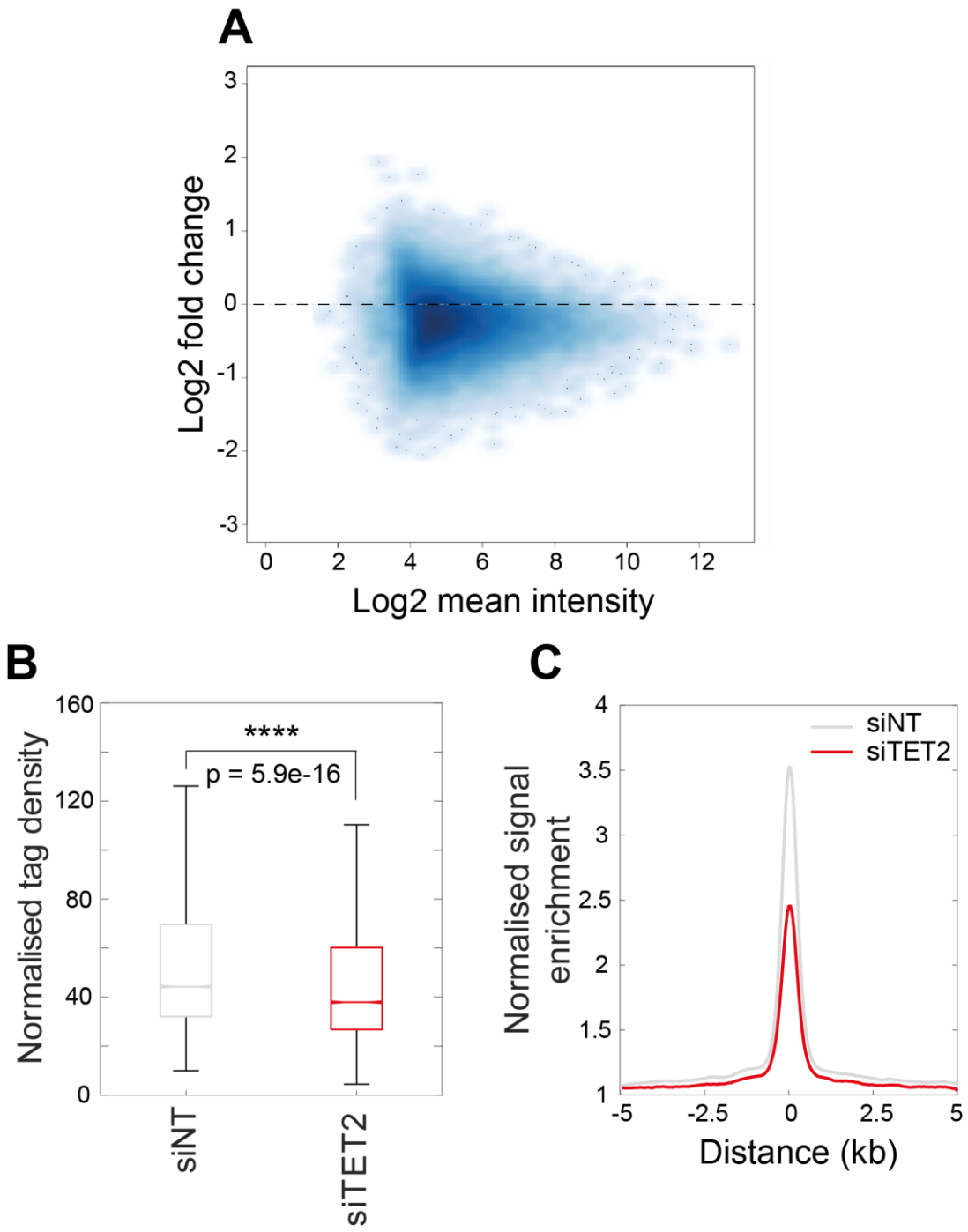


Figure 4.9. TET2 knockdown depletes global ER binding.
A) MA plot showing log2 fold change in ER binding in control versus TET2 knockdown conditions against the log2 mean intensity of ChIP-seq signal for all ER sites (20,386 peaks). B) Boxplot showing the normalised tag density of ER ChIP-seq signal in control (non-targeting control siRNA, siNT) and TET2 knockdown (siTET2) conditions within all ER peaks. **** = p ≤ 0.0001. C) Average plot showing normalised signal enrichment of ER ChIP-seq under control (siNT) or TET2 knockdown (siTET2) conditions within all ER peaks. TET2 knockdown was performed for 72 hours.

However, it remains possible that the effects of TET2 loss on ER binding may also be attributable to changes in TET2-mediated regulation of DNA modifications at ER sites. As described in Chapter 1, 5mC is generally considered to repress interactions of many DNA-binding proteins and transcription factors with their DNA motifs (Yin et al. 2017). As TET2 has been suggested to protect enhancers from aberrant methylation (Jeong et al. 2014; Rasmussen et al. 2015), loss of TET2 might result in an accumulation of 5mC at ER sites which could potentially destabilise ER binding, or the binding of other ER cofactors which might have a knock-on effect on the complex. Consistent with this, Rasmussen et al. (2019) showed that TET2 appears to mediate demethylation of enhancers in haematopoietic cells, and based on observations of increased chromatin accessibility, and the motifs consequently exposed at these sites, suggested this facilitates recruitment of lineage-specific transcription factors. However, 5mC may not solely destabilise protein-DNA interactions, and indeed some transcription factors may rather bind preferentially at motifs containing 5mC (Yin et al. 2017). It is hypothetically possible that increased binding or stabilisation of additional proteins near ER sites could also have a disruptive effect on the ER complex, and hence ER binding. The potential effects of DNA modifications on the binding of transcription factors are not confined to 5mC, as TET-oxidised 5mC intermediates 5hmC and 5fC have also each been shown to attract the binding of distinct "readers" in a sequence context-dependent manner (Iurlaro et al. 2013; Spruijt et al. 2013; Song & Pfeifer, 2016). Although speculative, it is possible that changes in the levels of these marks as a result of TET2 loss could also alter the binding of certain ER complex proteins. The reduced ER binding in response to TET2 loss could therefore be contributed to by changes in the dynamic regulation of 5mC and its oxidised forms at ER sites, a certain profile of which may be required for the assembly of a stable ER complex. Further investigations, including examination of DNA modification dynamics at these sites, will be required to assess this.

4.2.6 Acute ER depletion reduces TET2 binding at a subset of sites

As previously described, TET2 appears to lack a sequence-targeted DNA binding domain (Ko et al. 2013). It thus remains poorly understood how it is recruited to specific genomic regions. Wang et al. (2018) demonstrated a link between the ER complex and TET2 recruitment by showing that disruption of the ER complex with tamoxifen treatment over 72 hours reduces TET2-chromatin interactions. However, as TET2 is an ER target gene,

this was associated with significant depletion of TET2 protein levels. Although this result clearly demonstrates a role for ER in regulating TET2 chromatin occupancy, this approach renders it difficult to decouple the effects of TET2 protein-level regulation versus direct stabilisation on chromatin through ER. Furthermore, although ER-chromatin associations are decreased by acute (45 minutes) tamoxifen treatment, as shown by Hurtado et al. (2011), increased durations of tamoxifen treatment (6–24 hours) result in significant upregulation of global ER protein levels (Papachristou et al. 2018), and the effect of tamoxifen treatment on ER binding at later timepoints, such as at the 72 hour timepoint measured by Wang et al. (2018), is unclear. To assess a role for ER in directly stabilising TET2 on chromatin, it would therefore be useful to decouple depletion of ER-chromatin interactions from dramatic changes in TET2 protein levels.

The aim of this experiment was to deplete genome-wide ER binding to assess whether ER directly contributes to stabilisation of TET2 on chromatin, but to simultaneously minimise any effect on TET2 protein levels that would confound interpretation of the results. To this end, cells were treated with the SERD fulvestrant for an acute duration of three hours, with the assumption that promoting active degradation of existing ER protein would provide rapid depletion of ER-chromatin interactions whilst minimising protein-level changes in target genes, including TET2. Targeted proteomics (PRM) was then used to measure the levels of ER and TET2 in response to fulvestrant treatment. Within this timeframe, although only one ER peptide was detected due to technical limitations, a significant drop in ER total protein levels was observed using PRM (Figure 4.10), and ChIP-qPCR at several key ER regulatory sites confirmed effective depletion of ER-chromatin interactions (Figure 4.11). Illustrating the sensitivity of TET2 as an ER target gene, despite the short treatment duration, fulvestrant produced a modest depletion of TET2 total protein levels, assessed through measurement of five different TET2 peptides. However, this depletion was significant for only one of the peptides measured (corresponding to a ~20% reduction compared to control), with changes in levels of the other four TET2 peptides remaining non-significant (Figure 4.10).

ER depletion through acute fulvestrant treatment resulted in significantly reduced TET2 binding at a distinct subset of TET2 binding events (1,810 out of 20,599 peaks, ~9%), as assessed using DiffBind (Stark & Brown 2011) (Figure 4.12A). This indicates that the

modest depletion of global TET2 protein levels observed using PRM was not reflected by a global drop in TET2 chromatin occupancy. In contrast, TET2 binding remained unchanged at the majority (~90%) of sites, which could imply a particularly important function for ER in recruiting TET2 at the depleted subset of sites. When investigated more closely, TET2 "lost" sites corresponded to ER/TET2 shared sites with higher ER binding intensity and significantly greater frequencies of ER motifs, and motifs for key ER cofactors FOXA1 and GATA3, than the TET2 binding sites that did not change in response to treatment (Figure 4.12B and C). Furthermore, genes in proximity to these lost sites included key ER target genes that are also repressed in response to TET2 knockdown, including *PGR*, *CCND1*, *XBP1*, and members of the EGR and E2F protein families. Examples of significantly depleted TET2 binding sites in proximity to the *PGR* and *XBP1* genes are shown in Figure 4.13. Taken together, these results demonstrate that the sites where TET2 targeting is disrupted are the ER binding sites that are most robustly detected using ChIP-seq. This indicates these peaks as the ER sites that occur most frequently in the cell population, and likely those where ER target gene expression is immediately and directly modulated by the complex.

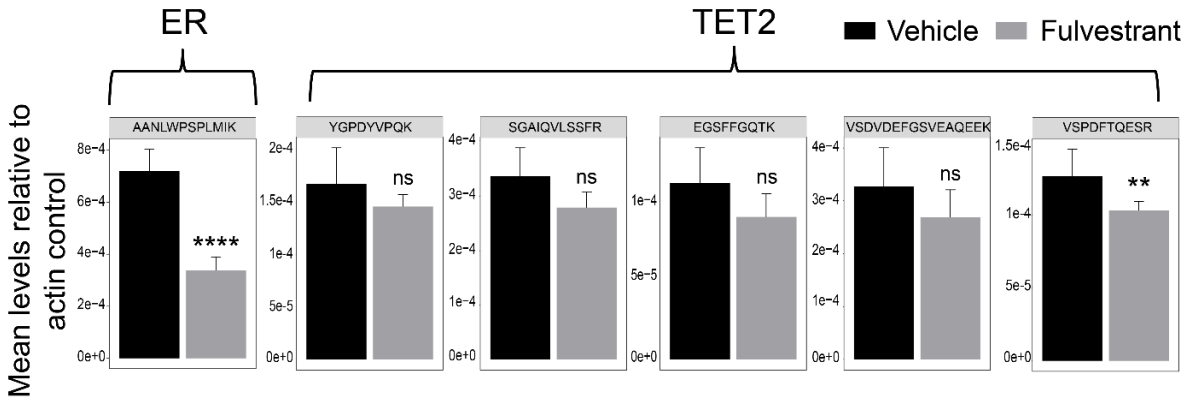


Figure 4.10. The effect of fulvestrant treatment on ER and TET2 total protein levels. PRM targeted proteomics was used to measure levels of ER- and TET2-specific peptides in MCF7 cells. Barplots depict levels of unique ER (left) or TET2 (right) peptides in response to vehicle (ethanol, 3 hours) or fulvestrant (100 nM, 3 hours). Results represent mean ± SD peptide levels relative to an actin control (n=4). ** = $p \leq 0.01$, **** = $p \leq 0.0001$.

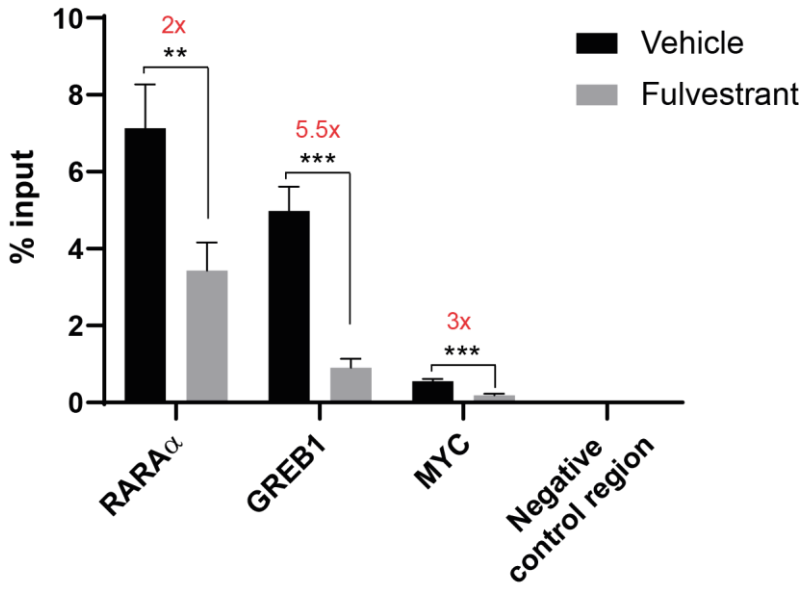


Figure 4.11. Fulvestrant treatment reduces ER chromatin occupancy.
ChIP-qPCR results showing reduction of ER-chromatin interactions in response to treatment with vehicle (ethanol, 3 hours) or fulvestrant (100 nM, 3 hours) at several key ER binding sites (error bars indicate mean ± SD, n=4). * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

It is of note that TET2 occupancy was significantly affected only at a subset of sites, despite the fact that (as shown in Chapter 3) almost all TET2 sites are co-occupied by ER in these cells. Whether ER is important for TET2 recruitment at the remainder of sites, or whether other mechanisms might target TET2 to these regions is unclear. Though a significant and robust depletion of total ER levels was observed (Figure 4.10), and ER binding was dramatically depleted at the key sites tested (Figure 4.11), it is possible that many sites retain levels of ER occupancy after 3 hours fulvestrant treatment that are sufficient for TET2 binding to remain unaffected. Longer durations of fulvestrant treatment might be required to disrupt TET2 binding at these remaining ER regions, at which point reduced TET2 expression due to ER depletion may become the dominant factor affecting TET2 occupancy. Having uncovered the sensitivity of TET2 as an ER target gene, it is clear that decoupling TET2-chromatin interactions from depletion of TET2 protein levels is challenging using this particular approach. Nevertheless, these investigations reinforce the association between ER and TET2, and demonstrate that, irrespective of whether through protein-level regulation or direct stabilisation on chromatin, ER has a key role in mediating TET2-chromatin interactions at direct ER binding sites.

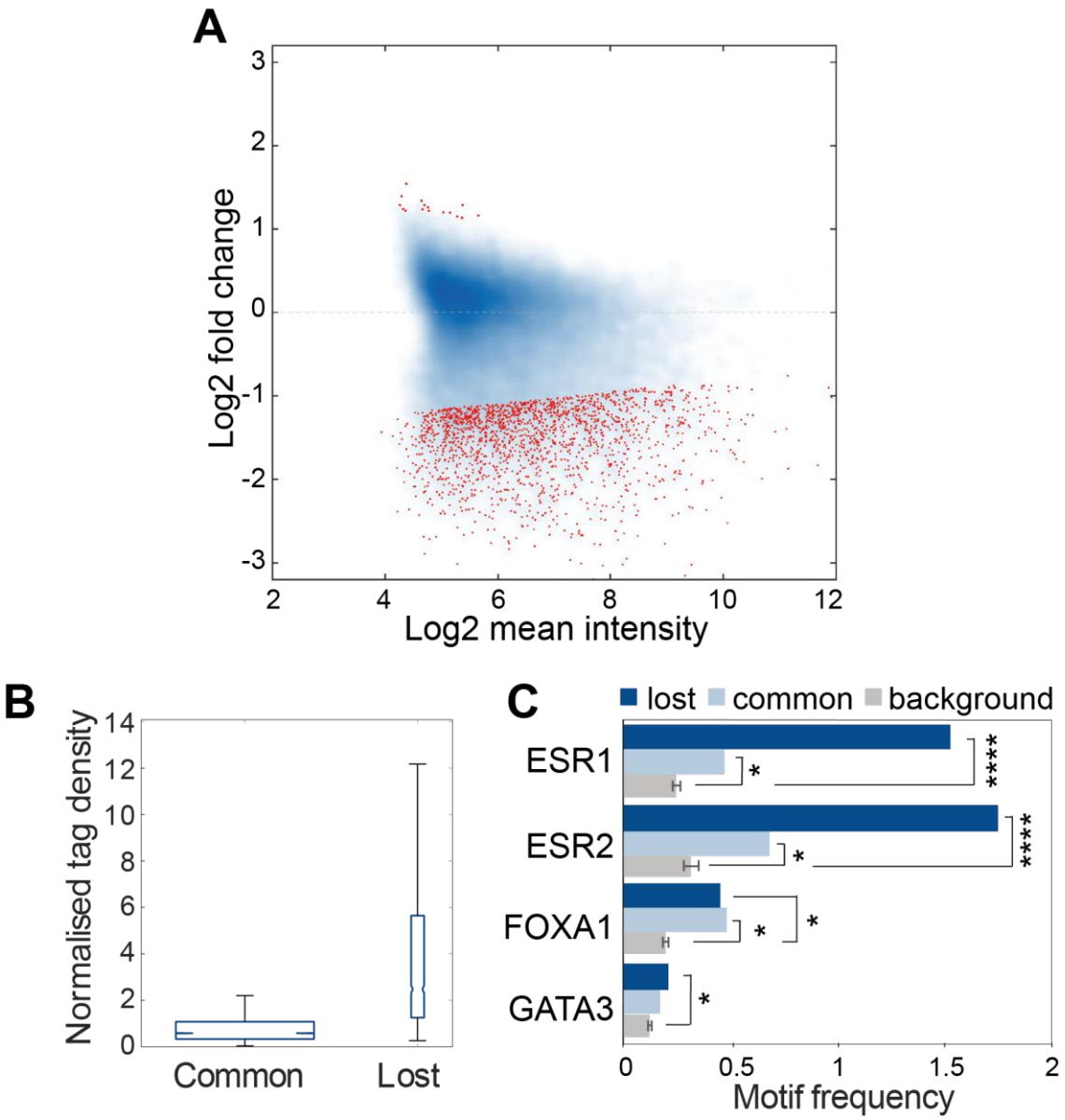


Figure 4.12. Acute ER depletion reduces TET2 binding at a subset of sites.

A) MA plot showing log2 fold change in TET2 binding in response to fulvestrant treatment (100 nM, 3 hours) against the log2 mean intensity of TET2 ChIP-seq signal for all TET2 sites (20,599 peaks). “Lost” sites (n = 1,810) and “gained” sites (n = 64) according to DiffBind analysis ($p \leq 0.05$) are highlighted in red. B) Boxplot showing the normalised tag density of ER ChIP-seq signal at unchanging (“common”) (n = 18,725) and “lost” (n = 1,810) TET2 sites in response to fulvestrant treatment. C) Barplot showing motif frequency (number of motifs divided by the total number of peaks in each category) of ER, FOXA1 and GATA3 motifs for lost, common and background sites. Background values were obtained using random open chromatin regions from an MCF7 MNase dataset (EBI Array Express E-MTAB-1958) and are expressed as the average ± SD of two separate background values calculated matched to the number of sites in the lost and common cohorts. Significance against background is indicated * = $p \leq 0.05$, ** = $p \leq 0.01$, **** = $p \leq 0.0001$.

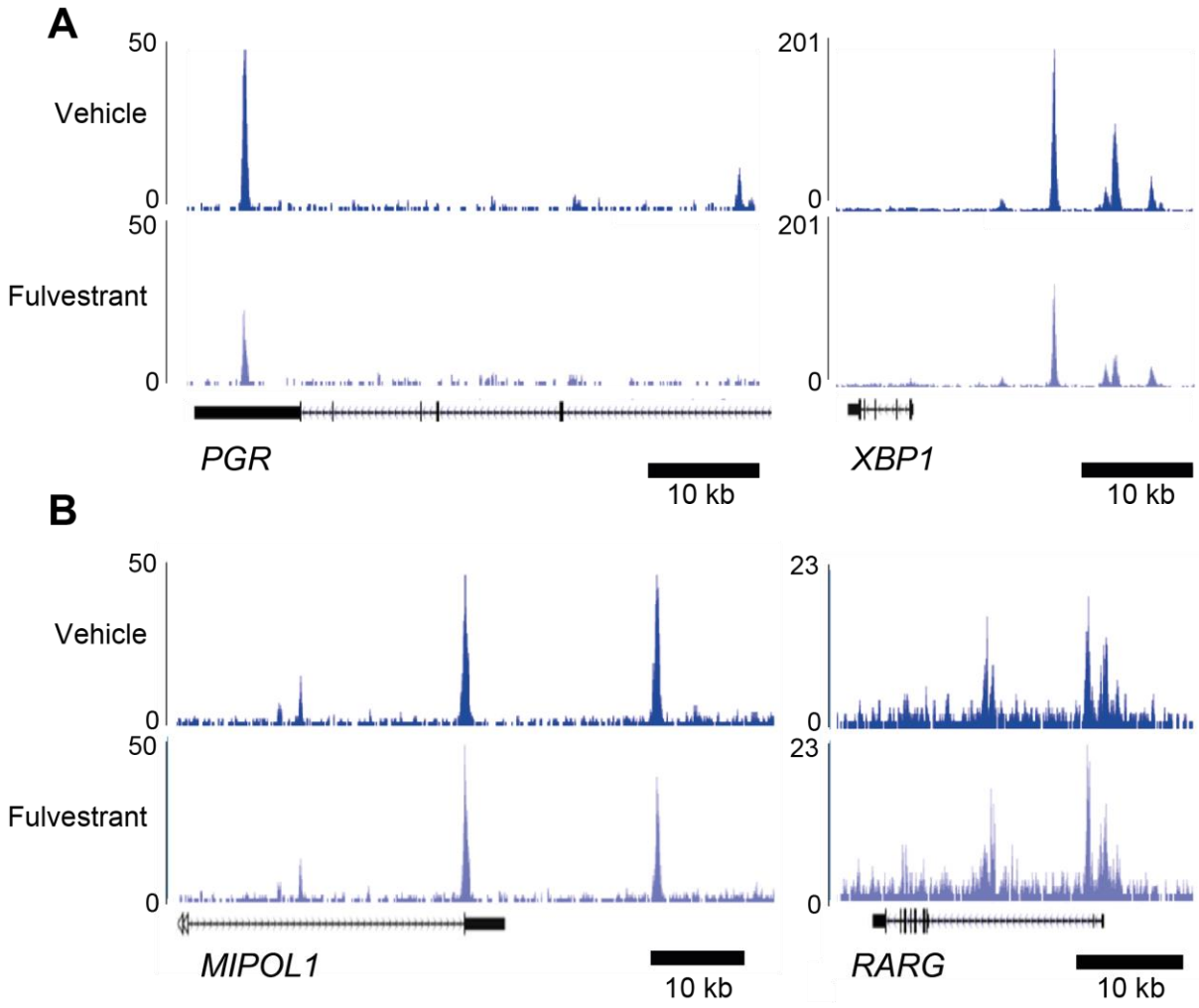


Figure 4.13. Fulvestrant-mediated depletion of TET2 binding at ER target genes *PGR* and *XBP1*.

A) UCSC genome browser tracks showing TET2 binding in response to treatment with vehicle (ethanol, 3 hours) or fulvestrant (100 nM, 3 hours) at significantly ($p \leq 0.05$) depleted sites according to Diffbind analysis, within 50 kb of the TSS of two key ER target genes (*PGR* and *XBP1*). B) UCSC genome browser tracks showing TET2 binding in response to treatment with vehicle (ethanol, 3 hours) or fulvestrant (100 nM, 3 hours) at unchanged sites according to Diffbind analysis. Tracks represent merged data from four biological replicates. For the gene schematics below each track, lines indicate introns, boxes indicate exons, and arrowheads indicate the direction of transcription.

4.3 Summary

The work in this chapter has shown that TET2 knockdown modulates gene expression in MCF7 cells, and that the TET2 gene regulatory programme includes a large proportion of genes also co-regulated by both GATA3 and ER. The effects of TET2 silencing included reduced activation of gene pathways related to cell proliferation, and consistent with this, MCF7 cell growth was inhibited by TET2 knockdown. Furthermore, a subset of the genes repressed by loss of TET2 coincide with those repressed in the therapeutic setting, as demonstrated by comparison with gene expression data from tamoxifen-treated MCF7 cells.

Whilst Chapter 3 demonstrated that ER and TET2 operate as part of the same complex and co-localise at chromatin, work in this chapter built on these findings by investigating the interdependence of these two factors in their ability to bind their target sites. It appears that TET2 has a role in stabilisation of the ER complex at chromatin, as ER-chromatin interactions were depleted, if modestly, in a global manner by TET2 knockdown. This is consistent with observations of the role of other oestrogen-regulated ER cofactors (such as FOXA1, GREB1 and NCOA3) in ER-chromatin interactions (Hurtado et al. 2011; Mohammed et al. 2013; Tikkanen et al. 2000), and likely contributes to the influence of TET2 loss on the expression of ER-regulated genes. Nevertheless, the precise mechanism behind the potential TET2-mediated stabilisation of the ER complex remains unclear, and aspects of this response, for example the apparent loss of ER binding at sites that may not be co-occupied by TET2, merit further investigation.

Reinforcing the association between ER and TET2, it was also shown that ER depletion through fulvestrant treatment affects TET2-chromatin interactions. The function of TET2 as an ER target gene renders it difficult to decipher the full extent of the role of ER in targeting TET2 to chromatin at all ER/TET2 shared sites. Further investigations would be required to decipher whether this effect might be a result of direct or indirect tethering of TET2 through the ER complex, or whether TET2 occupancy mainly reflects its protein-level control by ER. It is possible that both these scenarios operate. Nevertheless, this result indicates an important role for ER in stabilising TET2 at chromatin as part of the ER complex. Coupled with the RNA-seq data demonstrating crossover between ER and

TET2-regulated genes, the observations of the interdependence of TET2 and ER chromatin interactions paint a picture of a firmly interconnected transcriptional network, where disruption of either of these proteins perpetuates destabilisation of the complex as a whole, with associated effects on gene expression.

Of the TET2 gene regulatory programme, it was the genes repressed by TET2 knockdown that displayed the strongest association with proliferation. This could imply that TET2 acts as an activator in the ER complex, contributing to oestrogen-driven cell growth. Nevertheless, TET2 knockdown resulted in significant upregulation of an equal number of genes to those that were repressed. It is therefore possible that TET2 may act as a transcriptional repressor at some sites, though these may be less directly related to the ER gene programme. Further work is needed to determine the precise mechanisms of TET2 transcriptional regulation, and to unravel any aspects of context-dependency at different sites. One proposed mechanism of TET2 transcriptional control is through regulation of DNA modifications, and the potential relevance of this process at ER/TET2 sites is investigated in Chapter 5.

Overall, these investigations into of the effects of acute TET2 disruption have yielded important insights regarding the role of TET2 in ER signalling. A different type of investigation that could help to further establish the role of TET2 in the ER complex would be to use stable *TET2* knockouts to examine the effects of chronic TET2 loss on ER/GATA3 gene expression and ER binding profiles in breast cancer cells. As a continuation of the work detailed in this thesis, characterisation of in-house generated *TET2* knockout MCF7 cells is currently underway, and will provide a useful tool for further investigations into TET2-mediated gene expression, and the functional effects of TET2 loss in ER+ cells.

For the work in this chapter, full proteome analysis was performed by the Proteomics Core Facility, with further bioinformatic analysis performed by Dr Kamal Kishore of the Bioinformatics Core Facility. PRM analysis was performed by Dr Carmen Gonzalez Tejedo of the Proteomics Core Facility. Sequencing was performed by the Genomics Core Facility, and RNA-seq and ChIP-seq data processing and analysis were performed by Dr Igor Chernukhin.

Chapter 5

TET2-mediated regulation of DNA modifications in ER+ breast cancer cell lines

5.1 Introduction

The work in Chapters 3 and 4 showed that TET2 is a key component of the ER complex, with TET2 binding events a near-total subset of ER sites in several ER+ breast cancer models, and TET2 itself subject to expression-level regulation by ER. RNA-seq demonstrated that TET2 also appears to have a role in the ER gene regulatory programme, and TET2 silencing abrogates growth in ER+ MCF7 cells. Whilst these knockdown studies showed that TET2 appears to stabilise ER on chromatin, and this likely contributes to its apparent influence on ER-mediated gene expression, the precise mechanisms behind this remain unclear. Moreover, TET2 possesses intrinsic enzymatic functions that warrant investigation with respect to its role in the ER complex.

As described in Chapter 1, TET2 is an Fe(II)/α-ketoglutarate-dependent dioxygenase, and can oxidise methylated cytosines (5mC) to produce the additional DNA modifications 5hmC, 5fC and 5caC. By way of these steps, TET2 can also contribute to active DNA demethylation, as 5fC and 5caC can be removed through TDG-mediated base-excision repair and replaced with unmodified cytosine (He et al. 2011; Ito et al. 2011; Maiti & Drohat 2011; Tahiliani et al. 2009). Unlike 5mC, which is heritable across cell divisions (Bostick et al. 2007; Hermann et al. 2004), it is unclear whether oxidised forms of 5mC are maintained during DNA replication. Therefore, TET2-mediated conversion of 5mC to 5hmC, 5fC and 5caC likely also contributes to replication-dependent loss of these marks. As described in Chapter 1, whilst 5mC has an established role in transcriptional repression

in many contexts, a role for 5hmC in regulation of gene expression is also emerging, and, although less-investigated, some studies suggest 5fC and 5caC may also have specific functional roles related to transcription (Iurlaro et al. 2013; Iurlaro et al. 2016; Kellinger et al. 2012; Spruijt et al. 2013). Together, this implies diverse ways in which TET2 might act to modulate gene expression.

Genome-wide, the distribution of 5mC appears to be relatively stable across different mammalian cell types, but notably varies at gene regulatory regions (Ziller et al. 2013). Coupled with the finding that TET2 appears to preferentially localise at ER-regulated enhancers in ER+ breast cancer models (Chapter 3), and the broad evidence that differential DNA methylation patterns can predict breast cancer classification (and in particular ER status) (Fackler et al. 2011; Hill et al. 2011; Ronneberg et al. 2011), this led to the hypothesis that TET2 may contribute aspects of the ER+ phenotype through controlling DNA modifications at ER sites. The work in this chapter sought to investigate this, first through examination of the dynamics of these epigenetic marks in ER+ MCF7 cells in response to TET2 knockdown, and secondly by investigating these changes in the context of TET2-regulated gene expression.

5mC and 5hmC are the most abundant cytosine modifications, with 5mC present as approximately 4-5% of total cytosines, and 5hmC levels ranging between 0.03-0.7% of total cytosines, dependent on the cellular context (Globisch et al. 2010; Lister et al. 2009; Szwagierczak et al. 2010). The final products of TET metabolism, 5fC and 5caC, are present at levels at least 1000 times lower than 5hmC (Globisch et al. 2010; Ito et al. 2011), and are more difficult to reliably detect using current methods. The studies in this chapter therefore focused on measurements of 5mC and 5hmC in response to TET2 knockdown to assess the role of TET2 at ER sites, with a combination of techniques employed for this purpose. DNA mass spectrometry was used to measure global levels of these modified nucleotides in genomic DNA, quantifying 5mC/5hmC as a percentage of total cytosines. For site-specific analysis of 5mC and 5hmC, two sequencing-based techniques were employed, Methyl Midi-seq (MMS), and Reduced Representation Hydroxymethylation Profiling (RRHP). Both methods rely on restriction digest-based enrichment of areas of the genome containing high levels of modifiable CpG dinucleotides, and as both promoters and enhancers exhibit heightened GC content (ENCODE Project

Consortium 2012; Fenouil et al. 2012) these methods facilitate the investigation of gene regulatory regions. Furthermore, the restricted coverage of these techniques avoids the prohibitive sequencing costs of genome-wide methods and decreases the time required for data processing and analysis, allowing more rapid identification of potential features of interest.

The workflows for MMS and RRHP are described in Figures 5.1 and 5.2, respectively. MMS relies on a bisulfite sequencing approach, which, due to the chemical effect of bisulfite treatment on 5mC and 5hmC, does not distinguish these two marks. This assay therefore provides a combined measurement of total 5mC and 5hmC. In contrast, the RRHP method profiles 5hmC exclusively. The use of these two techniques in combination facilitates a more detailed insight into the levels of 5mC/5hmC at specific sites than the measurement of net 5mC/5hmC provided by the bisulfite-based MMS method alone. Using these methods described, the role of TET2 in regulating DNA modifications in ER+ cell lines was thus examined. Global measurements of 5mC/5hmC using DNA mass spectrometry facilitated initial, broad assessment of the effects of TET2 loss on 5mC and 5hmC, and subsequent genomic analysis allowed specific assessment of ER sites. Following this, TET2-mediated changes in DNA modifications at ER/TET2 sites were examined in the context of TET2-regulated gene expression.

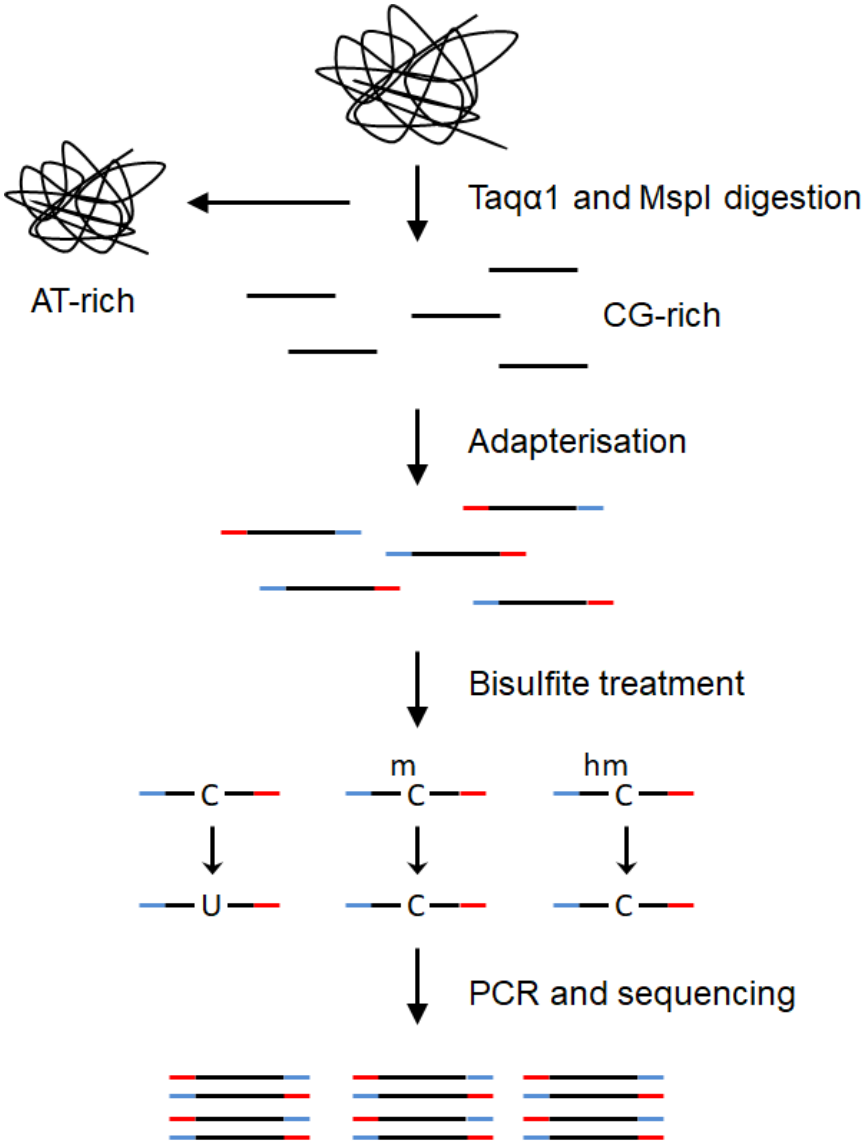


Figure 5.1. Methyl Midi-seq workflow.
Genomic DNA is digested with MspI (consensus site CCGG) and TaqαI (consensus site TCGA) to enrich for CG-rich areas of the genome. Fragments are then ligated to pre-annealed adapters containing 5mC instead of unmodified cytosine, such that their sequence will be preserved through subsequent bisulfite treatment. Bisulfite treatment results in deamination of unmodified cytosine to uracil, whilst 5mC and 5hmC are unaffected. Further modifications 5fC and 5caC, like unmodified cytosine, are also converted to uracil (not shown). Changes in the DNA sequence that depend on the methylation or hydroxymethylation status of individual cytosine residues are amplified by PCR and sequenced, providing single-nucleotide resolution of DNA modifications.

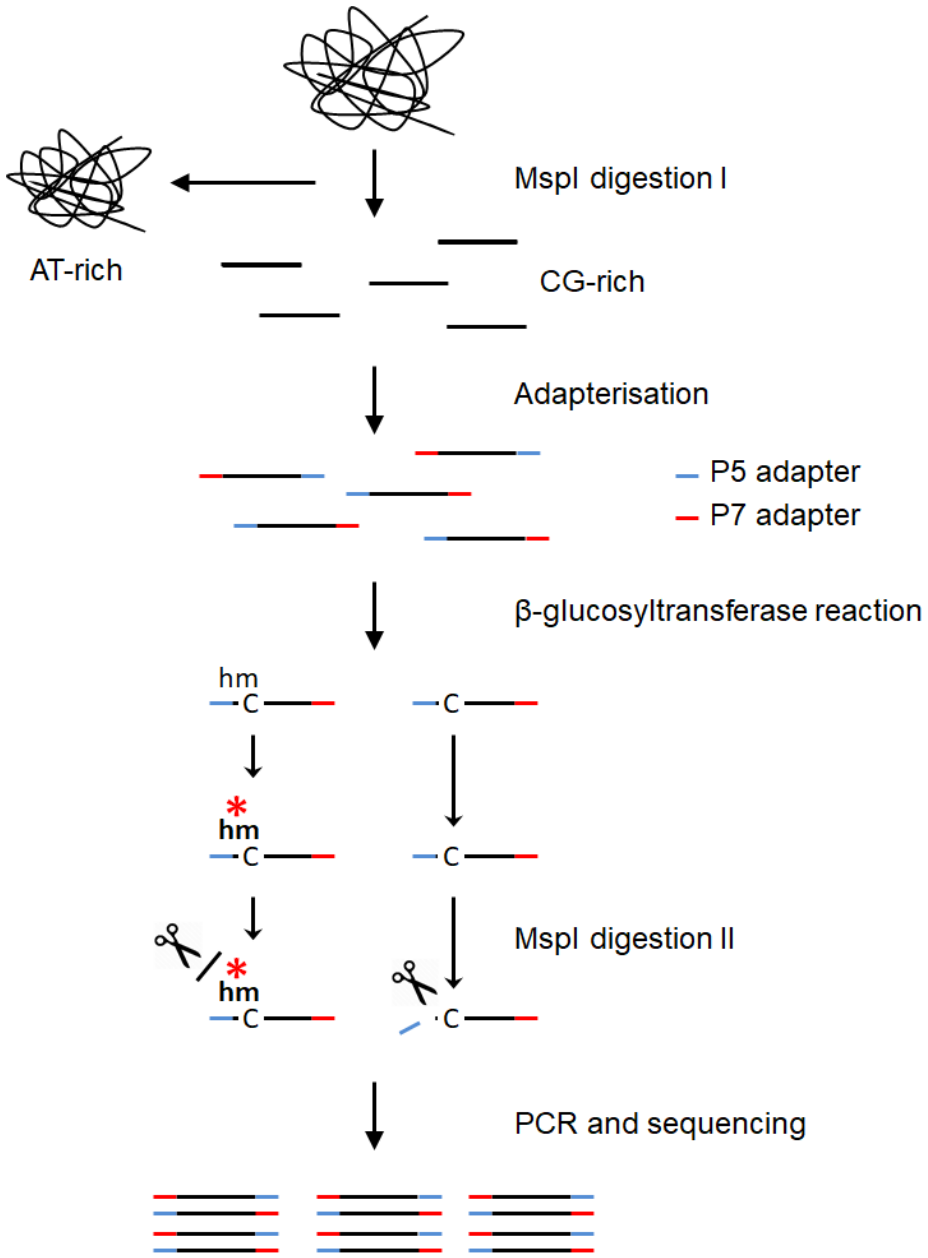


Figure 5.2. Reduced Representation Hydroxymethylation Profiling workflow. Genomic DNA is digested with MspI (consensus site CCGG) to enrich for CG-rich areas of the genome. Fragments are then ligated to pre-annealed adapters P5 and P7. Homo-adapterised fragments form an inhibitory stem-loop hairpin which will prevent further amplification, so only hetero-adapterised fragments are ultimately sequenced. The P5 adapter regenerates an MspI site at the adapter-fragment junction, whereas the P7 adapter does not. All ligated library fragments are next subject to a β-glucosyltransferase reaction, whereby 5hmC residues become glucosylated. Fragments presenting a glucosylated 5hmC at the CCGG at the P5-adapter junction are resistant to a second MspI digestion while fragments without 5hmC are cleaved, removing the P5 adapter. Only the fragments with both P5 and P7 adapters after the second MspI digestion, and thus possessing 5hmC at the P5-adapter junction, are amplified and sequenced. Adapted from Petterson et al. (2014).

5.2 Results

5.2.1 TET2 knockdown induces loss of global 5hmC, but has no effect on global 5mC levels cells according to mass spectrometry

To investigate the mechanistic role of TET2 in ER biology, mass spectrometry was used to measure total levels of 5mC and 5hmC in response to TET2 knockdown in MCF7 cells. TET2 protein levels were confirmed as robustly depleted at both 48 and 72 hours after TET2 knockdown using PRM, with four high confidence TET2 peptides detected, and all of these reduced by at least 85% compared to control (Figure 5.3B). As expected, the overall abundance of 5mC was much greater than that of 5hmC, with 5mC present as ~3% of total cytosines, and 5hmC present as ~0.02% of total cytosines, under control conditions. TET2 knockdown for 48 hours resulted in a dramatic decrease (~50%) in global levels of 5hmC in MCF7 cells, with this augmented to a ~75% drop after 96 hours of knockdown (Figure 5.3A), indicating that the effects of TET2 silencing on 5hmC are preserved across cell divisions (the doubling time of MCF7 cells is ~28 hours). Despite the robust drop in 5hmC in response to TET2 knockdown, no significant change in global 5mC was detected. Validation of these findings in additional ER+ cell lines T47D and ZR75-1 demonstrated similar effects (Figure 5.4), where, despite some variability in the measurements for the T47D cell line, an average drop of ~50% in global 5hmC levels was measured in response to TET2 knockdown for 48 hours, whilst global 5mC appeared largely unchanged (Figure 5.4A). PRM again confirmed effective depletion of TET2 protein levels in these cell lines (Figure 5.4B). Due to technical constraints imposed by the apparently lower levels of TET2 in T47D and ZR75-1 cells, only one of the five TET2 peptides probed using PRM was detected with high confidence. Nevertheless, qRT-PCR (Figure 5.4C) further confirmed expression-level depletion of TET2 in these cell lines.

The dramatic global decrease in 5hmC in response to TET2 loss is consistent with reports in the literature, with Hon et al. (2014) observing a global drop of 5hmC to ~10% of control levels in *TET2* knockout mESCs. This finding is backed up by studies in other cell lines, with the extent of 5hmC depletion varying in different systems, but generally appearing robust in response to TET2 loss (Li et al. 2011; Li et al. 2018). It is possible that compensatory activity of TET1 or TET3 could prevent an even more dramatic drop in

5hmC in response to TET2 silencing in MCF7 cells. However, as shown in Figure 5.5, neither TET1 nor TET3 mRNA levels demonstrated robust changes in response to TET2 knockdown according to RNA-seq in MCF7 cells, and indeed TET1 appeared very lowly expressed. Some redundancy between TET enzymes may still occur, and protein-level measurements and functional studies would be required to further validate these observations. Nevertheless, these findings suggest that of all three TET proteins, TET2 may exert the major control over 5hmC levels in these cells. These results indicate a key role for TET2 in the production and/or maintenance of 5hmC in ER+ breast cancer cells.

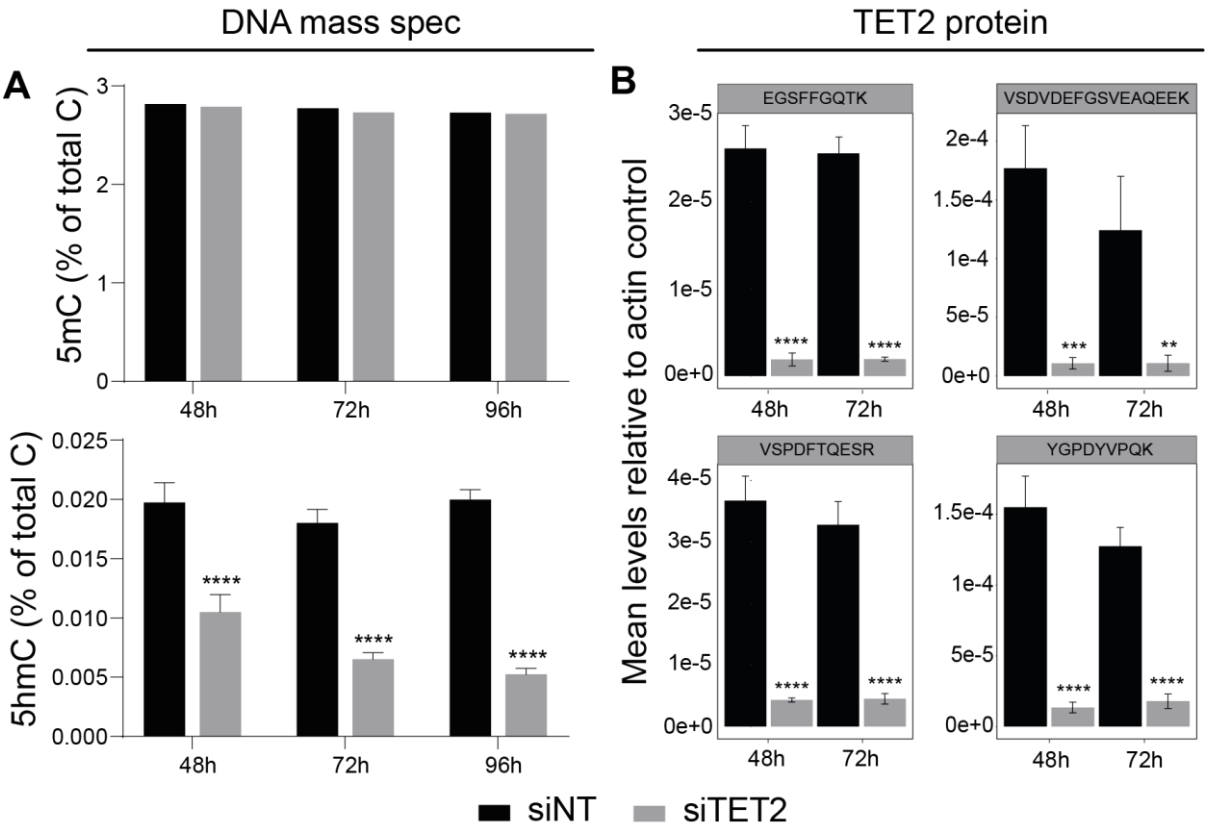


Figure 5.3. TET2 knockdown in MCF7 cells induces a global drop in 5hmC levels, but no change in overall 5mC levels.
A) Mass spectrometry was used to assess global levels of 5mC (top) or 5hmC (bottom) in DNA isolated from MCF7 cells treated with either non-targeting control siRNA (siNT) or siRNA targeting TET2 (siTET2) for various durations. Data represent mean ± SD (n≥4). Results are expressed as % of total cytosines. B) PRM confirmed effective TET2 knockdown in MCF7 cells. Barplots depict levels of different TET2 peptides in response to siNT or siTET2, with data representing mean ± SD peptide levels relative to an actin control (n=3). ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$.

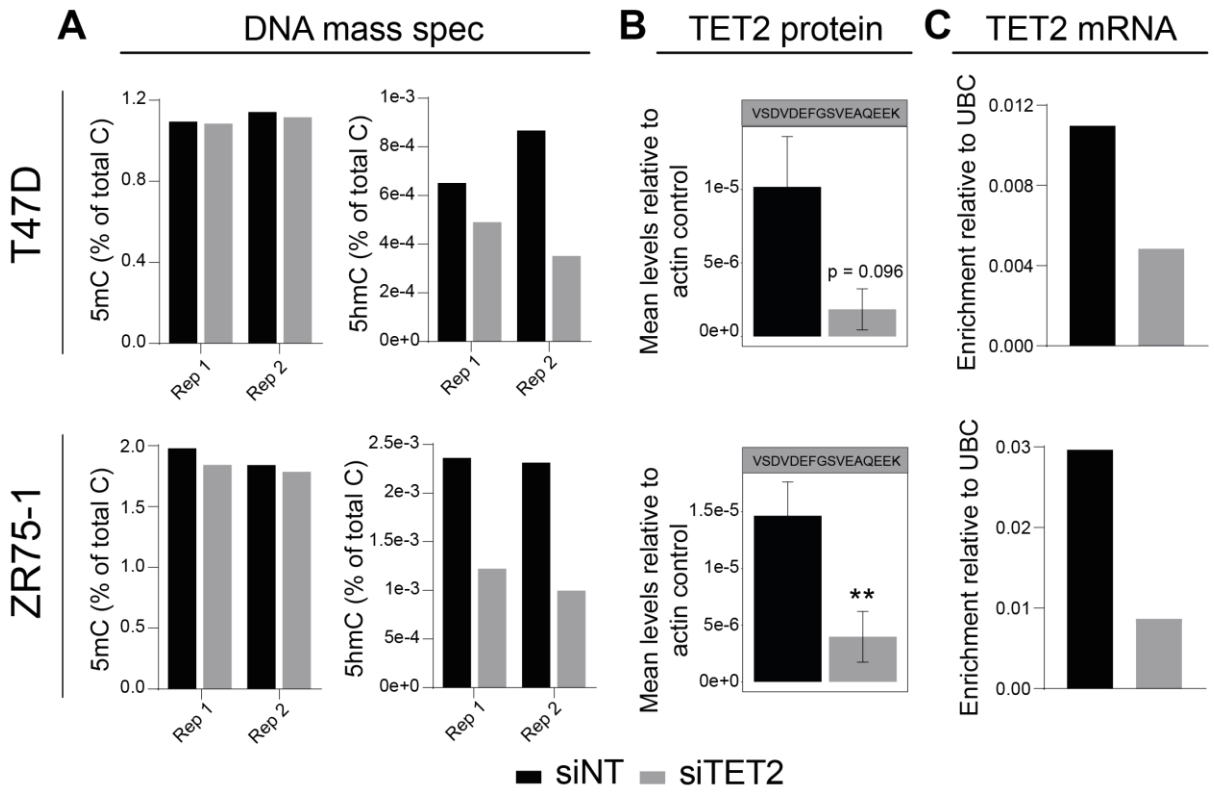


Figure 5.4. TET2 knockdown in T47D and ZR75-1 cells induces a global drop in 5hmC levels, but no change in overall 5mC levels.

A) Mass spectrometry was used to assess global levels of 5mC (left) or 5hmC (right) in DNA isolated from T47D cells and ZR75-1 cells treated for 72 hours with either non-targeting control siRNA (siNT) or siRNA targeting TET2 (siTET2). Measurements from two independent biological replicates are plotted separately. Results are expressed as % of total cytosines. B) PRM confirmed effective TET2 knockdown in T47D cells (top) and ZR75-1 cells (bottom) Barplots depict levels of the TET2 peptide VSDVDEFGSVEAQEEK in response to siNT or siTET2 treatment (72 hours), with data representing mean ± SD peptide levels relative to an actin control (n=3). ** = $p \leq 0.01$. C) To validate the PRM results, TET2 mRNA levels in T47D cells (top) and ZR75-1 cells (bottom) in response to siNT or siTET2 treatment (48 hours) were assessed using qRT-PCR. Data are expressed as enrichment relative to a housekeeping control gene (UBC) (n=1, average of three technical replicates).

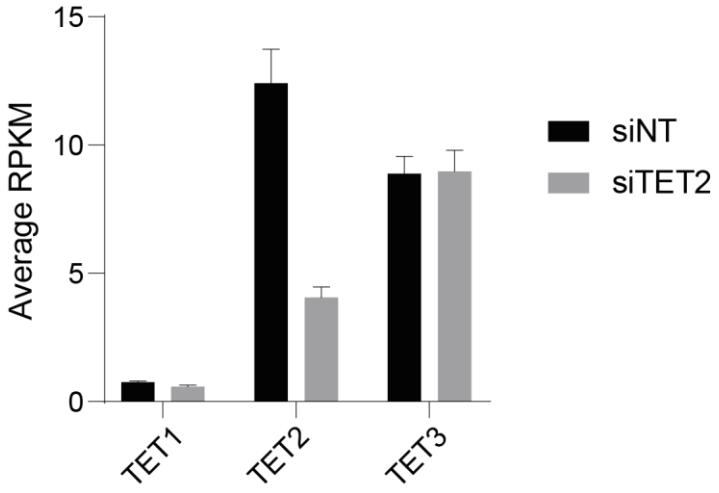


Figure 5.5. Expression of TET1, TET2 and TET3 mRNA in MCF7 cells in response to TET2 siRNA treatment.
Barplot shows the average Reads Per Kilobase of transcript per Million (RPKM) for TET1, TET2 and TET3 in MCF7 cells according to RNA-seq, in response to treatment with either non-targeting control siRNA (siNT) or siRNA targeting TET2 (siTET2) for 48 hours. Results represent mean ± SD (n=6).

The lack of a visible change in global 5mC levels in response to TET2 knockdown may be partly due to the bulk nature of these measurements. Specifically, whilst reports in the literature, cited above, show that TET2 loss results in global depletion of 5hmC, the picture appears to be more complex regarding loss of TET2 and 5mC. A recent study by López-Moyado et al. (2019) showed that TET2 knockout in a range of cell types causes localised increases in 5mC in active euchromatic regions, but a loss of 5mC in heterochromatic regions, implying that the effects of TET2 depletion on 5mC levels can vary dependent on chromatin state. It is therefore possible that the apparently stable readout of 5mC measured here masks site-specific variations. An alternative explanation could be that 5mC may only undergo regulation at discrete sites as a result of TET2 loss, consistent with dynamic modulation of this mark being observed mainly at selected regulatory regions across different cell types (Bird 2002; Ziller et al. 2013). If TET2 only impacts 5mC at a subset of sites, these changes may be subtle as a fraction of total 5mC, such that their detection is limited by the sensitivity of this technique. As the focus of this work is ER regulatory regions, subsequent investigations were aimed towards providing ER site-specific resolution of the effects of TET2 loss.

5.2.2 TET2 knockdown regulates DNA modifications at ER sites according to mass spectrometry

To focus on the effects of TET2 knockdown on 5mC and 5hmC at ER regulatory regions, ER ChIP was performed in asynchronous MCF7 cells and the DNA obtained was analysed using mass spectrometry. This showed that the global drop in 5hmC in response to TET2 loss was recapitulated at ER regulatory regions, where a ~50% loss of 5hmC was observed (Figure 5.6). Although very modest, there also appeared to be a concomitant increase in 5mC at ER sites in response to TET2 knockdown, indicating that TET2 loss may also impact 5mC levels at genomic regions bound by ER. Since only two biological replicates were conducted, statistical analysis was not performed on these results. This work therefore remains preliminary, and further replicates would be required to validate these findings. Nevertheless, this local increase in 5mC in response to TET2 knockdown would be consistent with studies indicating a role for TET2 in modulating 5mC specifically at enhancers (Hon et al. 2014; Wang et al. 2018), and potentially providing a protective effect against aberrant hypermethylation at these sites (Rasmussen et al. 2015).

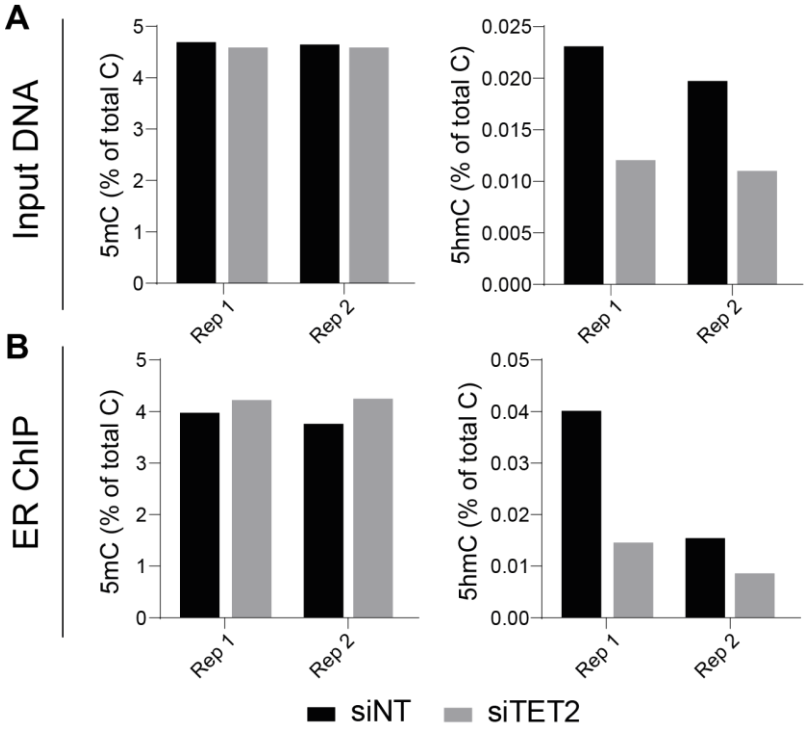


Figure 5.6. Analysis of ER ChIP DNA from MCF7 cells reveals changes in DNA modifications at ER sites in response to TET2 knockdown.
Mass spectrometry was used to measure 5mC or 5hmC levels in either A) input DNA or B) DNA from ER ChIP in MCF7 cells treated with either non-targeting control siRNA (siNT) or siRNA targeting TET2 (siTET2) for 72 hours. Measurements from two independent biological replicates are plotted separately. Results are expressed as % of total cytosines.

5.2.3 GATA3 knockdown reduces TET2 expression, but does not affect global 5mC or 5hmC according to mass spectrometry

As described in Chapters 3 and 4, GATA3 regulates the participation of TET2 in the ER complex. RNA-seq after GATA3 knockdown in MCF7 cells indicated that this is likely due to loss of TET2 expression, implicating TET2 as a GATA3 target gene. Having shown that TET2 knockdown results in changes in the levels of 5hmC (and, potentially, though more modestly, 5mC) at ER regulatory sites, it was reasoned that GATA3 might provide indirect modulation of these pathways. Mass spectrometry to measure global 5mC and 5hmC was therefore performed after GATA3 knockdown in MCF7 cells. As expected, GATA3 loss resulted in a robust reduction of TET2 protein levels according to PRM, with four high confidence TET2 peptides detected, and three of these depleted significantly by GATA3 knockdown (Figure 5.7C). However, this loss of TET2 did not translate into an effect on global 5mC or 5hmC (Figure 5.7A). There are several possible explanations for why TET2 depletion via GATA3 knockdown might not phenocopy direct TET2 silencing with regards to 5hmC changes. Firstly, the depletion of TET2 protein levels as a result of GATA3 knockdown is not as robust as the effect of TET2 knockdown itself. Whilst the TET2 peptides measured by PRM were reduced to less than 15% of control after direct TET2 silencing (Figure 5.3B), their levels ranged between 20-56% of control in response to GATA3 knockdown (Figure 5.7C). This indicates that there may be a threshold of TET2 loss required for disruption of 5hmC levels that is not achieved via GATA3 silencing. Another possibility is that any changes in 5mC/5hmC in response to GATA3 knockdown are both less dramatic than those induced by direct loss of TET2, and/or occur in a site-specific manner, and the bulk nature of these mass spectrometry measurements prevents resolution of these events. A more speculative explanation for the lack of 5hmC depletion is that GATA3 knockdown may simultaneously induce compensatory changes that stabilise 5hmC in response to reductions in TET2 protein levels. For example, increases in the levels or activity of TET1 or TET3 in response to TET2 loss could hypothetically facilitate increased 5hmC generation. However, as neither TET1 or TET3 mRNA levels were increased in response to GATA3 knockdown (data not shown), functional studies would be needed to explore this, or any other compensatory changes that might protect from loss of 5hmC in response to GATA3 knockdown. Off-target effects of the GATA3 siRNAs used could be a further confounding feature that may affect this facet of the response to GATA3 knockdown, although similarly, further in-depth investigations would be required to assess this.

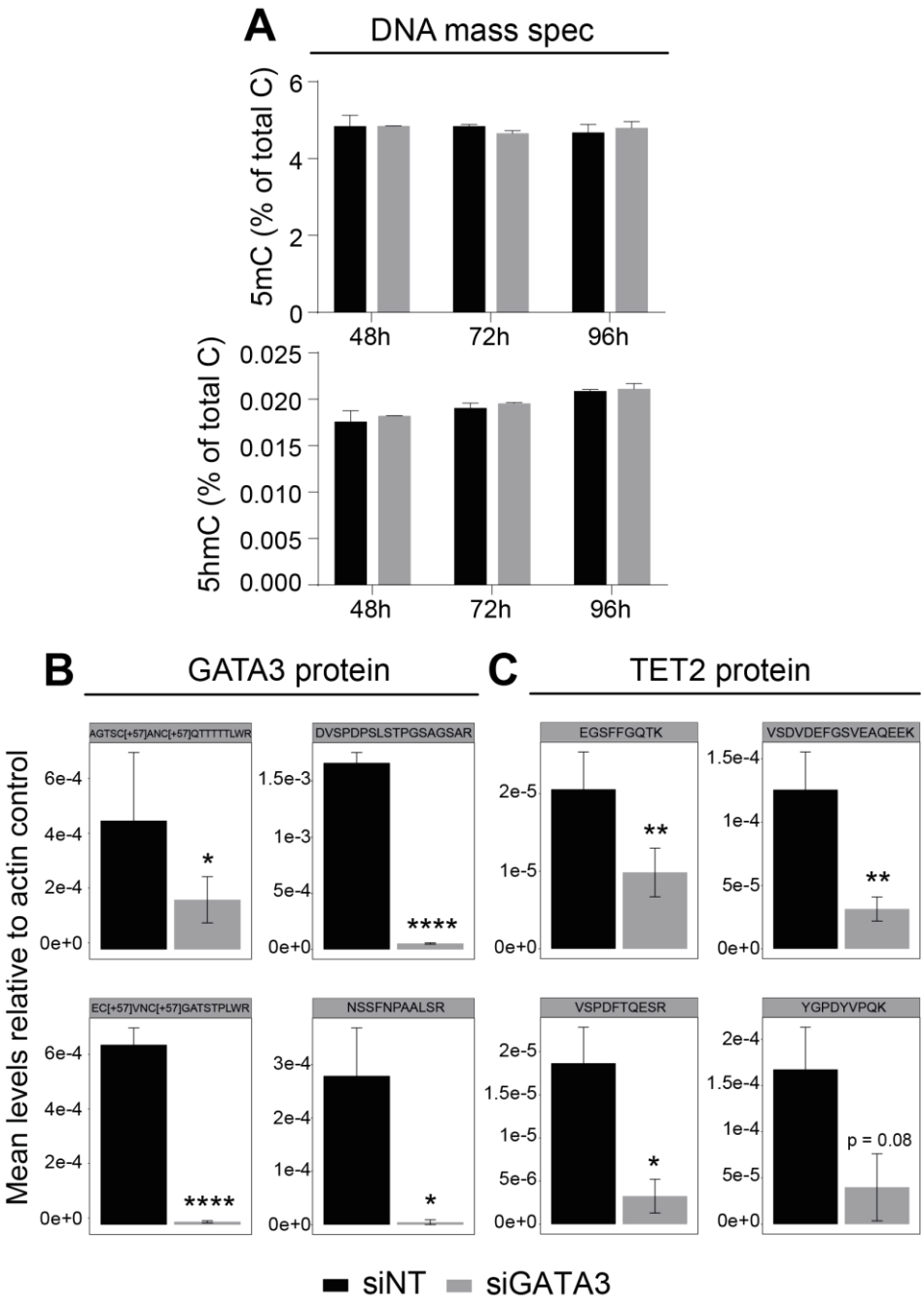


Figure 5.7. GATA3 knockdown results in depletion of TET2 protein levels, but no effect on global 5mC or 5hmC in MCF7 cells.

A) Mass spectrometry was used to assess global levels of 5mC (top) or 5hmC (bottom) in DNA isolated from MCF7 cells treated with either non-targeting control siRNA (siNT) or siRNA targeting GATA3 (siGATA3) for 72 hours. Measurements from two independent biological replicates are plotted separately. Data are expressed as a % of total cytosines. B) PRM confirmed effective GATA3 knockdown. Barplots depict levels of different GATA3 peptides in response to siNT or siGATA3 treatment (48 hours). C) PRM showing TET2 levels in response to GATA3 knockdown. Barplots depict levels of different TET2 peptides in response to siNT or siGATA3 treatment (48 hours). Results represent mean ± SD peptide levels relative to an actin control (n=3). ** = $p \leq 0.01$, **** = $p \leq 0.0001$.

5.2.4 Sequencing-based measurements of 5mC and 5hmC confirm that TET2 regulates DNA modifications at ER sites

To obtain greater resolution of TET2-mediated changes in DNA modifications at ER sites, sequencing-based measurements of 5mC and 5hmC were performed. A modified form of bisulfite sequencing, MMS, was used to profile total 5mC and 5hmC, and 5hmC was uniquely profiled using RRHP. These methods are outlined in Section 5.1 (Figures 5.1 and 5.2). Both these techniques use the methylation-insensitive enzyme MspI (consensus site CCGG) to enrich for regions with a high frequency of CpG dinucleotides, with MMS coverage enhanced through further digestion with TaqαI (consensus site TCGA). Both promoters and enhancers display heightened GC content (Consortium, 2012; Fenouil et al., 2012), although promoters overall more so due to the abundance of CGIs within these sequences. To confirm that MMS and RRHP would provide adequate representation of ER regulatory regions, bioinformatic analysis was performed to assess the number of ER peaks covered by theoretical MspI-generated fragments of sequencing-appropriate size. Using an MCF7 dataset of 29,149 ER peaks from Mohammed et al. (2015), and size-selecting for fragments between 100-400bp (as per the MMS and RRHP protocols) it was shown that 20% of ER peak regions overlapped with these MspI fragments, the majority (13%) of which were outside promoters. This indicated that these techniques could provide sufficient coverage of ER sites to allow investigation of 5mC/5hmC dynamics in response to TET2 loss, and ultimately, assessment of whether any observed changes might correlate with TET2-regulated gene expression.

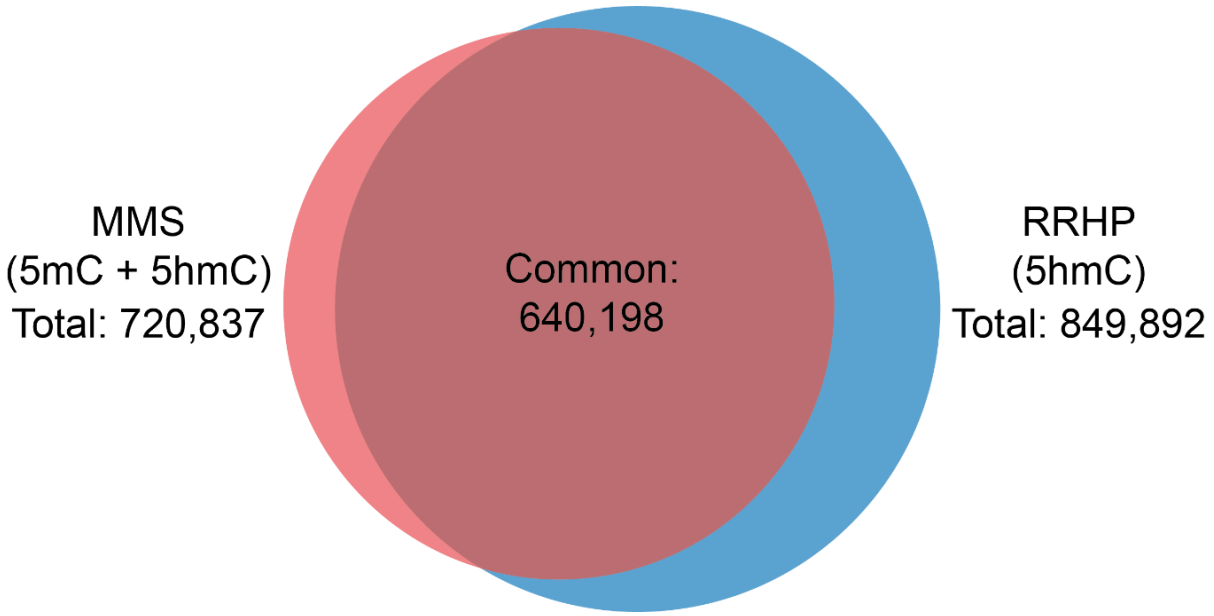


Figure 5.8. Venn diagram showing residue-for-residue overlap of modified bases detected using MMS and RRHP in MCF7 cells.
Numbers correspond to the number of sites exhibiting modifications in either MMS or RRHP in DNA obtained from MCF7 cells treated with non-targeting siRNA for 72 hours. MMS measures both 5mC and 5hmC, whilst RRHP detects 5hmC exclusively.

Using DNA obtained from MCF7 cells, 720,837 modifications were profiled using MMS, and 849,892 modifications were profiled using RRHP. The restriction digest methods of both techniques enrich for similar regions of the genome, whilst the bisulfite sequencing approach of MMS detects both 5mC and 5hmC. As a result, the modified sites profiled by each of these techniques demonstrated a high degree of residue-for-residue overlap (Figure 5.8). This indicates that the majority of sites detected using MMS are either stable 5hmC sites, or, more likely, present as 5mC modifications at some loci and 5hmC at others. Cell-to-cell variation in the type of modifications present at specific positions, and possibly allele-specific differences in 5mC/5hmC within the same cell (Song et al. 2019; Stern et al. 2017), may both account for this overlap. As TET-mediated oxidation of 5mC is a requirement for 5hmC generation, overall this implies that many of these sites are subject to dynamic regulation by TET proteins under normal conditions.

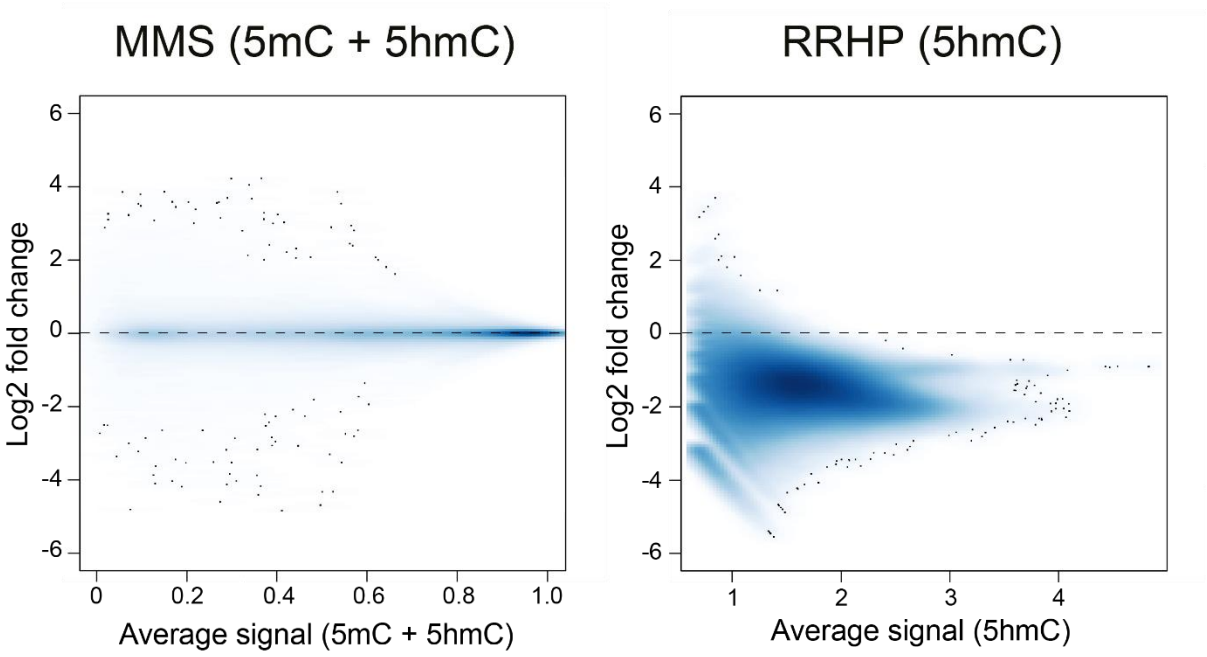


Figure 5.9. Residue-specific changes in 5mC and 5hmC in response to TET2 knockdown in MCF7 cells.
MA plots showing log2 fold change in 5mC + 5hmC (left, MMS readout) and 5hmC exclusively (right, RRHP readout) in control versus TET2 knockdown conditions. Each datapoint represents an individual 5mC or 5hmC residue. Knockdowns were performed for 72 hours.

The MA plots in Figure 5.9 show the global changes in response to TET2 knockdown reported by the MMS and RRHP assays. A dramatic drop in 5hmC was observed at the vast majority of sites where this modification was profiled using RRHP (Figure 5.9B), consistent with the mass spectrometry data (Figures 5.3 and 5.6). In contrast, most sites displayed no change in combined 5mC/5hmC levels (Figure 5.9A). The lack of a significant global shift at sites profiled by MMS (the vast majority of which, as implied in Figure 5.8, appear to be found both methylated and hydroxymethylated in the cell population) could suggest that, reflective of its higher genomic abundance, the frequency of 5mC at these sites is much higher than that of 5hmC. This would mean that the incidence of 5hmC measured by RRHP reflects a very small fraction of the total modifications detected by MMS for a specific residue, and the sensitivity of the MMS assay is not sufficient to distinguish loss of 5hmC at these sites from the stable overall 5mC signal. However, it is possible that the relative frequency 5mC and 5hmC could be more balanced at some of the regions profiled by MMS and RRHP, and Pastor et al. (2013) have suggested that at least at promoters, 5hmC may be enriched in absolute terms compared to 5mC. If this is

the case, the stability of the MMS signal in response to TET2 knockdown could imply that any loss of 5hmC equals a corresponding gain in 5mC at these individual sites. In either case, the clear depletion of 5hmC in response to TET2 loss, observed using both mass spectrometry and RRHP, is consistent with an important role for TET2 in dynamic regulation of the 5mC to 5hmC conversion in these cells.

Focusing on ER and TET2 co-bound genomic loci, 29% of total ER sites and 24% of ER/TET2 shared sites were found to contain at least one 5mC/5hmC residue as measured by MMS, whilst 35% of total ER sites and 28% of ER/TET2 sites possessed at least one 5hmC modification according to RRHP. Figure 5.10 shows MMS (5mC + 5hmC) and RRHP (5hmC) readouts under control or TET2 knockdown conditions at all ER sites, and at the subset of ER sites co-occupied by TET2. Interestingly, the overall drop in 5hmC as measured by RRHP was the same at both groups of sites, and hence the magnitude of 5hmC loss appears not to rely on the presence of TET2. This is somewhat consistent with the findings of Rasmussen et al. (2019), who showed that most differentially methylated regions identified in *TET2* -/- mESCs did not demonstrate detectable TET2 binding. This could imply that the TET2 antibodies used for ChIP, both in the work reported here, and the Rasmussen et al. (2019) study, did not effectively capture all TET2 sites. Alternatively, TET2 may regulate DNA modifications at these alternative sites in the absence of robust and persistent binding. This could imply a “scanning” mechanism whereby TET2 exerts its catalytic effects rapidly and with minimal interaction with chromatin. Another explanation is that some of these DNA modification changes may occur as secondary events to TET2 depletion, pointing to an indirect role for TET2 in regulating 5mC/5hmC levels.

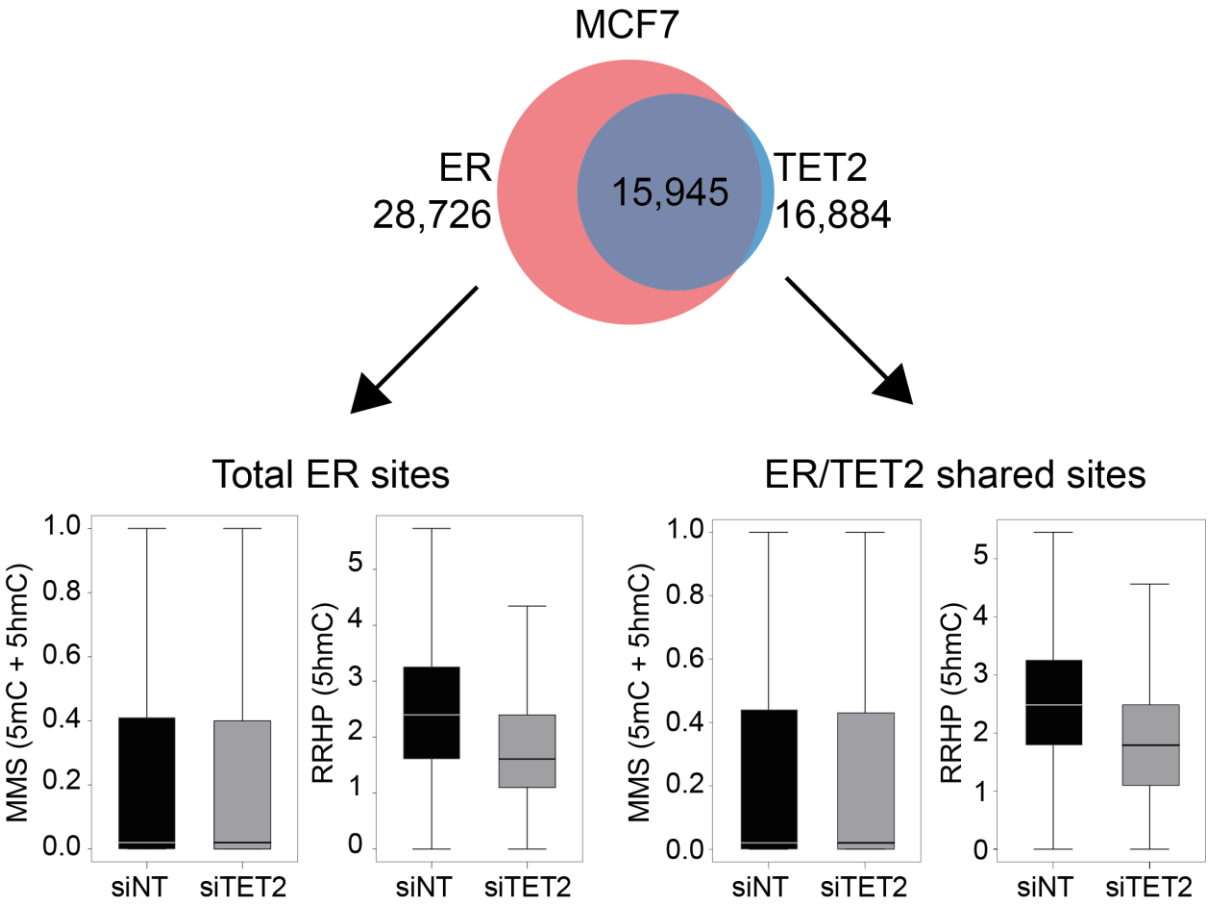


Figure 5.10. The effect of TET2 knockdown on DNA modifications at ER sites in MCF7 cells.
MMS signal (5mC + 5hmC) and RRHP signal (5hmC) was assessed at ER peak regions (left) or ER/TET2 overlapping peak regions (right) under control (siNT) or TET2 knockdown (siTET2) conditions. The total numbers of sites analysed within each category are: ER MMS sites = 8,463; ER RRHP sites = 10,104; ER/TET2 MMS sites = 3,762; ER/TET2 RRHP sites = 4,512.

5.2.5 5mC/5hmC changes reveal no direct correlation with TET2-regulated gene expression

To investigate a potential relationship between TET2-mediated regulation of DNA modifications at ER sites, and changes in the expression of nearby genes, the 500 most induced and 500 most repressed genes in response to TET2 silencing (sorted based on p-value, all $p \leq 0.05$) were selected. A further 500 genes with unchanged expression in response to TET2 knockdown were selected as controls, and the average MMS and RRHP readouts at ER/TET2 shared sites within ± 50 kb of the TSSs of these genes were examined (Figure 5.11).

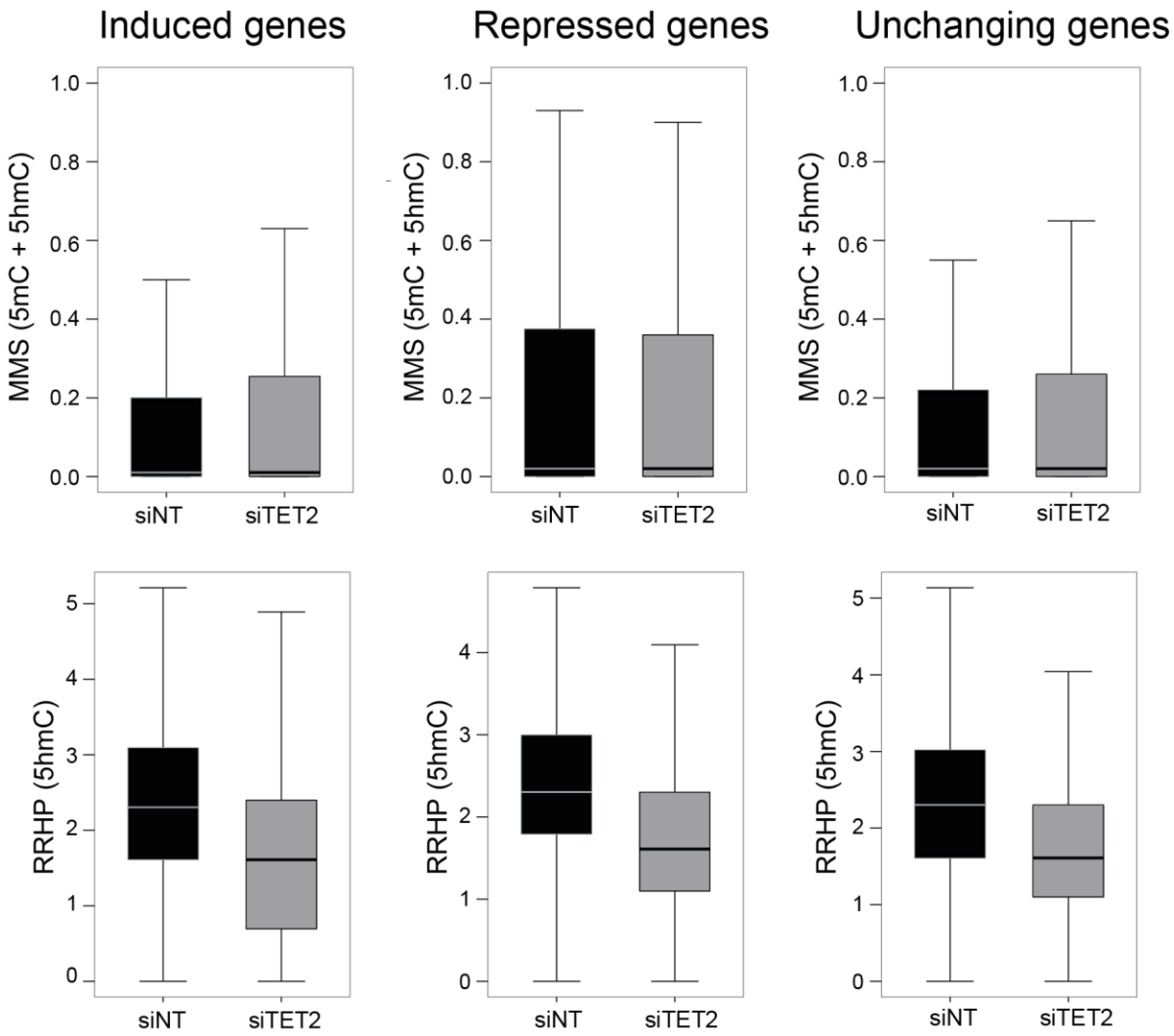


Figure 5.11. TET2 knockdown in MCF7 cells shows a uniform effect on DNA modifications at ER/TET2 shared sites in proximity to genes regulated by TET2 knockdown, and genes with unchanged expression.
MMS signal (5mC + 5hmC) and RRHP signal (5hmC) was assessed under (siNT) or TET2 knockdown (siTET2) conditions (72h) at ER/TET2 shared sites within ± 50 kb of the TSSs of the top 500 induced or repressed genes (sorted based on p-value, all $p \leq 0.05$) in response to TET2 knockdown in MCF7 cells. Unchanging genes (n = 500) were defined as those with $p > 0.1$ and log2 fold change < 0.05.

Subdivision of ER/TET2 sites according to the TET2-mediated regulation of nearby genes did not reveal differences in the MMS readouts at these different groups of regions. Specifically, the net levels of 5mC + 5hmC remained similar in magnitude and unchanged between control and TET2 knockdown at sites near induced, repressed and unchanging genes. In terms of 5hmC, whilst a robust depletion of 5hmC was measured across all three classes of sites in response to TET2 knockdown, reflecting the global measurements, no overall differences in the magnitude of this change were observed in proximity to induced, repressed and unchanging genes. Although, as shown in Chapter 3, both ER and TET2 express a preference for binding to enhancers in these cells, they also interact at promoters, and it has been suggested that 5mC and 5hmC may have distinct transcriptional contributions at these two classes of sites. However, separate examination of ER/TET2-occupied promoter regions (within 2 kb upstream and 1 kb downstream of the TSS) compared to enhancers did not distinguish the MMS and RRHP readouts nearby induced, repressed or unchanging genes from those at ER/TET2 shared sites as a whole (data not shown). Genes regulated in response to TET2 loss therefore showed no clear bias towards differential regulation of DNA modifications at ER/TET2 sites. As these analyses, although more targeted than global measurements, nevertheless involve averaged signal over groups of sites, it remains possible that divergent effects in the direction or magnitude of any changes in DNA modifications might be observed at individual sites. In addition, as this investigation focuses on the most significantly regulated genes, there may be differences in the extent to which these may be affected by changes in DNA modifications, compared to genes undergoing subtler regulation in response to TET2 loss. TET2-mediated regulation of 5mC/5hmC may therefore provide a regulatory contribution at subsets of these genes, or in a subtler manner, in concert with other mechanisms.

Overall, these results imply that complex and potentially divergent mechanisms might be involved in regulation of these different groups of genes in response to TET2 loss, with a more granular relationship existing between TET2-regulated gene expression and changes in DNA modifications. As described in Chapter 1, whilst many studies have demonstrated a correlation between gene expression and 5mC/5hmC levels at certain genomic elements, this relationship appears to be non-uniform, with several mechanisms of context-dependency reported for these effects. For example, CpG content can define the transcriptional effects of methylation at different types of promoters. Specifically, whilst

methylation of high and intermediate CpG density promoters is associated with repression, methylation of low CpG density promoters appears to have less of an effect on gene expression (Marchal & Miotto 2015), and it is possible that a similar concept may apply to enhancers. Similarly, the existing levels of 5mC or 5hmC within different regions might influence both the impact of TET2 loss on the modifications at these sites, and any effects on the expression of associated genes. Additional variables could include the histone modifications present at these sites, as it has been shown in mESCs that the profiles of histone modifications at active versus bivalent promoters correlates with differential changes in gene expression in response to hypermethylation of these regions (Lu et al. 2014). Furthermore, dynamic interplay between cytosine modifications and transcription factors or chromatin remodellers (a variety of which participate in the ER complex) may also play a role in the relationship between 5mC/5hmC levels and gene expression. It is emerging that a range of proteins may be capable of interacting with 5mC and 5hmC, and that these may be capable of diverse transcriptional functions, further reinforcing that the profiles of these 5mC/5hmC readers at specific sites are likely to be important in dictating the effects of these epigenetic marks on gene expression (Iurlaro et al. 2013; Spruijt et al. 2013; Spruijt & Vermeulen 2014). The affinity of certain 5mC and 5hmC readers may also vary dependent on the DNA sequence surrounding these marks (Iurlaro et al. 2013), and it is also possible that differential post-translational modifications of 5mC/5hmC readers might affect their affinity for their sites of interest, resulting in different functional outcomes (Spruijt et al. 2013). Finally, several additional roles of TET2 were described in Chapter 1, for example TET-mediated stabilisation of the histone methyltransferase complex SET1/COMPASS (Deplus et al. 2013). Some of these functions may be catalytic-independent, and may also contribute to context-dependent effects of TET2 on gene expression, to fine-tune the role of TET2 at different sites (Lian et al. 2016). Taken together, these findings paint a potentially complex picture of the role of TET2 in transcriptional regulation.

5.3 Summary

The work in this chapter demonstrates that TET2 regulates DNA modifications in several ER+ breast cancer cell lines, and shows more specifically in MCF7 cells that this process occurs at ER binding sites. Given that ER sites mainly constitute distal regulatory elements, this is consistent with the finding that these are amongst the key regions where

DNA modifications appear to undergo TET2-mediated dynamic regulation (Ziller et al. 2013; Rasmussen et al. 2015; 2019; Wang et al. 2018).

Regarding the precise nature of these DNA modification changes, DNA mass spectrometry indicated a global depletion of 5hmC in response to TET2 loss, indicating a key role for TET2 in regulation of this mark in ER+ breast cancer cells. No effect was observed on global 5mC, and further sequencing-based measurements of 5mC and 5hmC using a bisulfite approach (MMS) revealed no overall changes in response to TET2 knockdown, both at the global level and at ER sites specifically. In the absence of a direct measure of 5mC, whether TET2 may prevent aberrant methylation at ER sites is not completely clear. The preliminary mass spectrometry data following ER ChIP (Figure 5.6) suggests that TET2 loss may indeed result in hypermethylation at some ER sites, although further replicates are required to validate this. Nevertheless, it appears clear that TET2 has an important role in the conversion of of 5mC to 5hmC at these regions, with the possibility that 5hmC then remains stable and functional in its own right. Investigating whether TET2 activity at some ER sites might also involve demethylation in the classic sense (i.e. restoration of unmodified cytosine) may require further site-specific examination of these results. Whilst this work was in progress, a study was published by Wang et al. (2018) examining the effects of stable TET2 knockout in MCF7 cells. Consistent with measurements obtained using the acute TET2 knockdown system employed in this thesis, Wang et al. (2018) observed no global change in overall bisulfite sequencing signal in TET2 knockouts compared to control cells. However, an increase in bisulfite signal, suggesting an overall gain in 5mC, was observed at a subset of ER sites. This increase in 5mC occurred to different extents at different clusters of ER sites, with no changes seen at the promoters analysed, suggesting that some ER enhancers, but not all, may undergo demethylation as a result of TET2 activity. These observations validate the association between ER and TET2 shown in this work, and also suggest that *bona fide* TET2-mediated demethylation may occur at some ER sites in response to TET2 loss. The measurements of 5hmC provided in this chapter build on existing findings by demonstrating direct evidence of TET2-mediated regulation of 5hmC levels at ER enhancers in breast cancer cells.

In terms of the functional effects of these changes in DNA modifications, subdividing differentially hydroxymethylated (and by inference, potentially differentially methylated) sites into those near genes repressed, induced or unchanged in response to TET2 loss revealed no direct correlation between DNA modification changes and gene regulation. Therefore, while it is possible that TET2-mediated regulation of DNA modifications has a role in controlling gene expression at ER sites, work in this chapter indicates that this appears to be a more nuanced relationship than that of 5mC/5hmC gain/loss equating to transcriptional control in a particular direction. In the above-cited study from Wang et al. (2018), comparison of TET2 knockout cells and wild-type controls revealed that the observed changes in DNA methylation at subsets of ER sites were associated with changes in oestrogen-regulated expression of nearby genes. It is therefore possible that more detailed sub-classification of regulated genes, or inspection of individual loci, may reveal an association between the dynamics of TET2-regulated gene expression and changes in 5mC/5hmC in the acute knockdown system employed in this work.

Of further interest, TET2 loss appeared to coincide with 5hmC depletion throughout the regions profiled using MMS and RRHP, irrespective of the presence of TET2. This implies either that not all TET2-chromatin interactions have been captured using ChIP, or that loss of 5hmC can somehow occur secondary to TET2 depletion. This could also imply that the role of TET2 at ER/TET2 shared sites is unique to its contribution at other sites, related to an increased occupancy time of TET2 on chromatin, reflected in TET2 being capturable by ChIP at these regions. In addition, catalytic-independent functions, in combination with the observed regulation of 5mC/5hmC, may help to distinguish the potential effects of TET2 at ER sites from its putative role at other genomic regions. Selective inactivation of the catalytic domain of TET2 (e.g. through CRISPR) would facilitate decoupling of TET2 enzymatic function from its role in the ER complex and could provide the first step towards answering this complex question.

For the work in this chapter, DNA mass spectrometry analysis was performed by Dr Shiqing Mao. PRM analysis was performed by Dr Carmen Gonzalez Tejedo of the Proteomics Core Facility, and ChIP-seq data processing and analysis was performed by Dr Igor Chernukhin. MMS and RRHP data was obtained through Zymo Research (Irvine, California), and was analysed by Dr Kamal Kishore of the Bioinformatics Core Facility.

Chapter 6

Discussion

6.1 GATA3, TET2 and the ER complex

In Chapter 3, quantitiative proteomics was used to assess the composition of the ER complex in response to knockdown of GATA3, with the aim of investigating the contribution of this key ER-associated transcription factor to ER signalling. As a result of GATA3 silencing, several proteins with links to DNA methylation exhibited altered interactions with the ER complex. Specifically, the dioxygenase enzyme TET2 was depleted from the ER complex, whilst the transcription factors LHX4 and ZBTB34 were enriched. TET2 is thought to be capable of mediating DNA demethylation through its iterative oxidiation of 5mC to 5hmC, 5fC and 5caC (He et al. 2011; Ito et al. 2011; Tahiliani et al. 2009), as, unlike 5mC, which is heritable, these marks are likely lost through replication-dependent dilution (Hackett et al. 2013; Inoue & Zhang, 2011). Moreover, 5fC and 5caC may additionally be actively removed via base excision repair (He et al. 2011; Shen et al. 2013; Song et al. 2013). Meanwhile, LHX4 and ZBTB34 are transcription factors thought to be capable of binding to methylated DNA (Filion et al. 2006; Qi et al. 2006; Yin et al. 2017). The altered levels of these proteins in response to GATA3 knockdown implies that GATA3 might provide a functional link to writing and reading of DNA methylation as part of the ER complex. Specifically, these results might suggest that changes in DNA methylation through loss of TET2 could cause altered LHX4 and ZBTB34 binding to ER regions. Subsequent work (discussed further in sections 6.2 and 6.3 below) explored the relationship between TET2, the ER complex, and ER-regulated gene expression in more detail. TET2- and GATA3-mediated regulation of DNA modifications in ER+ breast cancer cells was investigated in Chapter 5, with TET2 demonstrating a role in regulating 5hmC levels at ER sites.

In Chapter 3, TET2-ER interactions were characterised using both RIME and ChIP-seq, and TET2 was shown to co-localise with ER on chromatin at a subset of sites in four different ER+ breast cancer models (two cell lines and two PDXs). Of note regarding the profile of TET2 genomic binding, it was shown in Chapter 5 that TET2 silencing in MCF7 cells results in loss of 5hmC at sites where no TET2 binding is detected. The presence of indirect TET2-regulated 5hmC events could be due to the nature of TET2 action, for example a sampling or scanning mechanism whereby TET2 does not fully interact with chromatin when oxidising cytosine modifications. A further, more speculative explanation could be that some of these apparently indirect TET2-mediated 5hmC events are the result of chromatin looping. Alternatively, the TET2-regulated 5hmC changes in the absence of TET2 binding could imply that not all TET2 sites are captured by ChIP-seq in these cells, despite the antibody chosen for these studies emerging as the most robust and specific TET2 antibody tested in both ChIP and RIME. The profile of TET2 binding in the models examined might also insinuate the presence of additional, uncaptured TET2 binding events, as TET2 binding was seen to correlate very closely with that of ER, but only at a subset of sites. Additionally, in all four models, very few ER-independent TET2 sites were observed. Whilst the latter may be a biological property of TET2 action in these cells, consistent for a stabilising role of ER in TET2-chromatin interactions (as discussed in Chapter 4), this finding merits further investigation. Any hypothetically uncaptured TET2 binding sites, whether ER co-occupied or ER-independent, likely correspond to lower-affinity binding sites that exhibit weaker TET2 enrichment and hence cannot be captured with the antibody used. One method to circumvent this, and to additionally test the specificity of the TET2 antibody, could be to express the TET2 protein tagged with an epitope that can be robustly and specifically detected in ChIP, for example FLAG or HA. Expression of this tagged TET2, in either in untreated cells or after knockdown of the endogenous protein, could provide validation of the existing TET2 sites and might point to additional sites where TET2 is found at low levels but cannot be captured in its endogenous form. This could also enable the detection of ER-independent TET2 binding events, which might aid in dissecting the precise role of TET2 in the ER complex as opposed to at other loci.

TET2 is an ER target gene, induced in response to oestrogen and repressed by tamoxifen in MCF7 cells (Papachristou et al. 2018; Wang et al. 2018). The ER qPLEX-RIME experiment in Chapter 3, coupled with the subsequent RNA-seq investigations, showed

that TET2 is also a GATA3 target gene. As GATA3 forms part of the ER complex at many sites, and is linked to ER function, the expression-level control of TET2 by GATA3 may be a correlate of its regulation by the ER complex as a whole. This would reinforce that the basis of the functional interplay between these three proteins is to sustain the transcriptional activity of the ER complex. Having detected TET2 as a GATA3-regulated ER interactor, an area that remains to be explored is the precise relationship between GATA3 and TET2 genomic binding. Notably, like TET2, GATA3 also associates with a distinct subset of ER sites (Theodorou et al. 2013). It would therefore be informative to assess how GATA3 and TET2 genomic binding compares and to what extent they participate in the same complex. Indeed, if the presence of GATA3 additionally helps to stabilise the TET2 protein in the ER complex, GATA3/TET2 overlap could account for the specific profile of TET2 binding.

Future work to establish the consequences of GATA3 and TET2 loss from the ER complex could also include investigating the factors enriched in response to GATA3 knockdown. For example, as previously mentioned, the recruitment of LHX4 and ZBTB34 to the ER complex may partly reflect their upregulation in response to GATA3 knockdown, but could also be a result of aberrant methylation, due to the concurrent loss of TET2, that attracts the binding of these proteins to ER sites. Indeed, TET2 has been suggested to protect against hypermethylation of enhancers (Jeong et al. 2014; Rasmussen et al. 2015),. Increased methylation of certain ER sites is observed in stable TET2 knockout cells (Wang et al. 2018), and preliminary mass spectrometry data in this thesis points to increased overall methylation at ER sites in response to acute TET2 knockdown, assessed through bulk analysis of ER ChIP DNA (Chapter 5). Nevertheless, an examination of DNA modifications at ER sites specifically in response to GATA3 knockdown remains a requirement of these investigations. Indeed, in notable contrast to the robust changes in 5hmC levels induced by TET2 silencing, loss of TET2 secondary to GATA3 knockdown had no visible effects on global 5hmC levels (Chapter 5). This implies a more nuanced relationship between GATA3 knockdown and TET2-mediated control of DNA modifications, and suggests that loss of GATA3 does not phenocopy loss of TET2. An opposing suggestion might therefore be that the binding of LHX4 and ZBTB34, or the additional factor GREB1L, also enriched in response to GATA3 knockdown, could protect against loss of 5hmC and/or aberrant methylation as a result of TET2 depletion through

this pathway. Further functional studies, including a detailed examination of the profiles of 5mC and 5hmC at ER sites after GATA3 knockdown, would be required to assess this.

In contrast to LHX4 and ZBTB34, which are transcription factors, the specific function of the third factor enriched in the ER complex in response to GATA3 loss, GREB1L, is currently unknown. However, Brophy et al. (2017) suggested this protein may have a role in renal development through interactions with RAR, implying it may be a general nuclear receptor cofactor. Indeed, GREB1L bears 51% protein sequence identity with the ER cofactor GREB1 using Protein BLAST (https://blast.ncbi.nlm.nih.gov) (Mohammed et al. 2013). Notably, proteins of this clade may also have a link to DNA modifications, with structural and comparative genomics analyses by Iyer et al. (2013) suggesting GREB1-like proteins may be capable of glycosylating 5hmC. However, no direct evidence of this has yet been shown. ChIP-seq for GREB1L, LHX4 and ZBTB34 under normal conditions and in response to GATA3 knockdown in ER+ breast cancer cells could help establish whether they are enriched at other regions prior to GATA3 loss, before being recruited to ER sites. In this way, a better understanding of the transcriptional role of these factors could shed further light on the contribution of GATA3 to the ER complex, and identify any potential consequences of TET2 loss at ER binding sites.

6.2 TET2 as an ER cofactor and potential prognostic factor in breast cancer

TET2 is an ER target gene in MCF7 cells (Papachristou et al. 2018; Wang et al. 2018), and ChIP-seq presented in Chapter 3 demonstrated robust ER binding sites 20-30 kb upstream of the TET2 promoter in these cells. The observation that ZR75-1 cells and the two PDXs examined also possess similarly-located ER binding sites suggests TET2 also functions as an ER target gene in these models. Work in Chapter 4 further reinforced the association between ER and TET2 by demonstrating that global ER binding is affected by TET2 silencing, and that this is associated with dysregulated expression of several ER and GATA3 target genes and reduced growth of ER+ MCF7 cells. These findings are consistent with those of Wang et al. (2018), who showed that in stable TET2 knockout MCF7 cells, oestrogen-induced recruitment of ER to chromatin, and oestrogen-driven cell

growth, are both attenuated compared to wild-type controls. The identification of TET2 as an ER cofactor therefore suggests the existence of a feed-forward loop in which co-regulators such as TET2 are expressed to support the transcriptional activity of the ER complex.

Consistent with the notion of TET2 as a key component of a functional ER complex, higher TET2 expression predicts improved relapse-free survival in ER+ breast cancer (Chapter 3), indicating TET2 as a potential prognostic factor. Having demonstrated the role of TET2 in ER-driven proliferation in MCF7 cells, the exploration of other malignant phenotypes, for example migration and invasion, could be of use in further characterising the role of TET2 in ER biology. Given the hypothesis that TET2 contributes to functionality of the ER complex, this protein could also have a role in the response to antihormone therapy. As ER+ breast cancer relapse is often associated with drug resistance, it would be pertinent to explore the effect of TET2 loss or overexpression in the context of long-term tamoxifen treatment. Transwell or scratch-wound assays could be used to examine proliferative, migratory and invasive phenotypes in cell lines, and mammary intraductal injection-based approaches using the MIND technique (Sflomos et al. 2016) could be used to investigate the behaviour of TET2-overexpressing or knockout cells *in vivo*.

TET2 expression is repressed in response to tamoxifen treatment in MCF7 cells (Papachristou et al. 2018), and it is also possible that loss of TET2 from the ER complex might result in transcriptional changes that, in a long-term clinical setting, could eventually contribute to resistance. Interestingly, Takayama et al. (2015) have shown that the androgen antagonist bicalutamide differentially regulates TET2 expression in hormone-dependent versus hormone refractory prostate cancer cells. Specifically, TET2 expression was induced by bicalutamide in the sensitive setting but repressed in resistance, which led to speculation that TET2 may have a role in clinical complications associated with the withdrawal of anti-hormone treatment in prostate cancer. Although reported in a different cancer type, this provides further evidence of a connection between TET2 and nuclear receptor activity in hormone-dependent cancers, and suggests that this relationship may become dysfunctional in the resistant setting. It would therefore be of interest to explore how TET2 levels are regulated in the context of long-term tamoxifen treatment in breast cancer cell lines or patient-derived xenograft systems.

In terms of DNA modifications in cancer progression, global depletion of 5hmC is observed in several cancers, and is associated with loss of TET2 (Kudo et al. 2012; Lian et al. 2012; Takayama et al. 2015; Yang et al. 2013). Moreover, lower levels of 5hmC may be specifically associated with more aggressive phenotypes, as has been shown in prostate cancer (Takayama et al. 2015). Dysregulated control of DNA modifications in response to TET2 loss could therefore play a role in antihormone resistance. Indeed, differentially methylated genes have been identified in tamoxifen-resistant versus wild-type MCF7 cells, and this differential methylation was associated with changes in the expression of genes related to proliferative control (Fan et al. 2006; Stone et al. 2012). Profiling 5mC and 5hmC in hormone-sensitive versus hormone-resistant breast cancer cell lines could help to solidify the relationship between TET2 loss and breast cancer progression.

Regarding TET2 in the context of cancer progression, it may also be useful to consider the breast cancer models used in this work. The MCF7 cell line is derived from a metastatic pleural effusion, and genome-wide analysis of ER binding in MCF7 cells compared to clinical samples has suggested that this cell line more closely resembles a poor outcome/metastatic tumour than tumours of earlier stages, despite its responsiveness to tamoxifen (Ross-Innes et al. 2012). The observation that TET2 expression may be low in both this cell line (based on the issues detecting this protein using Western blot), and in similarly metastasis-derived T47D and ZR75-1 cells (based on the technical difficulties in PRM analysis of TET2 in these cell lines), further implies that TET2 may be lost during breast cancer progression. The PDX models employed in Chapter 3 demonstrated co-localisation of TET2 and ER on chromatin, providing evidence of TET2 activity in an additional breast cancer system. However, it may be of interest to assess TET2 expression directly in clinical samples, which could represent earlier stages of breast cancer. Method development is currently underway to measure ER and TET2 levels in various clinical samples using PRM, combined with global 5mC and 5hmC measurements using mass spectrometry. Comparing these features with the clinical status of these tumours could provide an indication of the role of TET2 in these different settings, and any association between ER levels, TET2 levels, and 5mC/5hmC.

6.3 Mechanisms linking TET2 and ER biology

To assess the role of TET2 in regulating DNA modifications, global levels of 5mC and 5hmC were examined in response to TET2 loss in MCF7, T47D and ZR75-1 cells using mass spectrometry. Consistent with observations in other cell lines (Hon et al. 2014; Li et al. 2011; Li et al. 2018), loss of TET2 resulted in a dramatic global drop in 5hmC, indicating a key role for TET2 in regulation of 5hmC in these models. No effect was observed on global 5mC levels in the three ER+ cell lines tested. This may be partly due to TET2 loss resulting in site-specific changes in 5mC that are averaged out using these global mass spectrometry measurements, resulting in a stable overall signal. Alternatively, TET2 knockdown may rather induce 5mC changes only at discrete sites, resulting in alterations in 5mC levels that are subtle relative to its global abundance. Detection of these changes might therefore be limited by the sensitivity of the mass spectrometry technique.

To examine the potential role of TET2 at specifically ER sites, 5mC and 5hmC were profiled using MMS and RRHP in ER+ MCF7 cells. Reflecting the global mass spectrometry measurements, 5hmC was robustly depleted at ER sites in response to TET2 loss. In contrast, the exact nature of any 5mC regulation in response to TET2 loss could not be precisely quantified. This is because the assay used to assess 5mC levels (MMS) provides a composite measure of both 5mC and 5hmC, rather an an isolated 5mC readout, and this overall signal remained unchanged across ER sites in control and TET2 knockdown conditions. It is possible this signal, averaged over groups of ER sites, masks subtly divergent site-specific effects. Mass spectrometry analysis of ER ChIP DNA, which provides a measurement of 5mC alone, showed that loss of TET2 may cause increased 5mC levels at overall ER regions, although this requires further replicates for validation (Chapter 5). Nevertheless, this could indicate a role for TET2 in either protection of ER sites from aberrant methylation by conversion of 5mC to 5hmC, and/or in triggering full demethylation to unmodified cytosine at these sites. Indeed, Wang et al. (2018) suggested that stable knockout of TET2 in MCF7 cells may mediate demethylation of subsets of ER enhancers. However, emerging evidence also suggests that 5hmC, rather than being a transient intermediate in an enzymatic process leading to DNA demethylation, is instead created and remains stable in genomic DNA for an extended period of time (Bachman et al. 2014). TET2 activity might therefore facilitate maintenance of 5hmC at ER sites, where it acts as a stable epigenetic mark in its own right.

Distinction between 5mC and 5hmC can be attained through oxidative bisulfite sequencing (OxBS-seq). This is achieved by comparing sequencing results obtained from traditional bisulfite sequencing (5mC + 5hmC) with those obtained from a bisulfite sequencing method incorporating an oxidation step, which allows 5hmC to be sequenced as thymine whilst 5mC is read as cytosine (Booth et al. 2013). Comparison of the resulting sequences allows distinction between 5mC and 5hmC. As the 5hmC profiling method used in this thesis (RRHP) relies on enrichment-based detection of 5hmC rather than detecting specific changes to the DNA sequence as a proxy for the presence of these modifications, further investigation would be required to establish the feasibility of 'subtracting' the RRHP 5hmC readout from the MMS readout in this way. However, to investigate the dynamics of 5mC alone, OxBS-seq could be used to isolate the 5mC readout from 5hmC and interrogate these marks at base-resolution. This could be of importance to these studies, as this may help provide a clearer indication of which functions of TET2 are most important ER sites; whether TET2 activity predominantly involves turnover of 5mC to stable 5hmC, or whether this enzyme may mediate DNA demethylation at ER sites. This could provide a more comprehensive view of the contribution of TET2 to transcriptional regulation at ER regulatory regions. Performing investigations such as these after GATA3 knockdown could also help delineate any relationship between the loss of GATA3 and DNA modifications at specific ER sites.

Despite a clear role for TET2 in regulating 5hmC in ER+ breast cancer cell lines, no direct correlation was observed between these changes at overall ER sites in response to TET2 loss, and the directionality of the expression changes of nearby TET2-regulated genes. The approach of grouping ER/TET2 sites into those near upregulated, downregulated and unchanging genes, although providing a broad overview, may average out any site-specific differences in 5mC/5hmC that could contribute to differential gene expression at different sites. In addition, several factors may contribute to a more nuanced relationship between gene expression and changes in DNA modifications. Some of these (such as CpG content of differentially methylated promoters, and differential assembly of chromatin remodellers and other transcriptional regulators at different sites) were discussed in Chapter 5. An alternative approach could be to first subclassify ER/TET2 sites based on their different properties, then examine any changes in DNA modifications at these sites in conjunction with changes in the expression of nearby genes. Specific properties of the ER/TET2 sites could then be taken into account when assessing the impact of any DNA

modification changes on gene expression. For example, categorising ER/TET2 sites based on their profile of histone modifications could provide an alternative starting point. The finding that increases in 5mC at promoters bearing H3K4me3 and/or downstream H3K79me2/3 (initiated or active promoters) and their associated enhancers is linked to gene repression, whilst increases in 5mC at promoters bearing H3K4me3 and H3K27ac (bivalent promoters) and their associated enhancers is linked to gene activation (Lu et al. 2014) implies that the combined effects of histone marks and changes in DNA modifications could be important in defining gene regulatory outcomes at different sites.

Additional factors to consider in the potentially nuanced relationship between 5mC/5hmC changes and gene expression could include the strength of ER/TET2 binding, and the profile of additional ER cofactors present at specific loci. For example, in Chapter 3 it was shown that ER interacts with both TET2, which may be involved in maintaining low 5mC at its target sites, and DNMT1, which methylates cytosines to produce 5mC. It is possible that DNMT1 and TET2 could participate in the ER complex to exert opposite effects on DNA methylation, and at sites where their levels or activity in the ER complex do not counter one another equally, the changes in DNA modifications might result in a fine-tuning of transcriptional control, at a subset of sites and in a locus-specific manner. Indeed, a study by Ariazi et al. (2017) suggested ER may direct methylation of certain target genes through an association with DNMTs. In addition, oestrogen-induced methylation and silencing of the *CYP1A1* promoter in ER+ MCF7 cells was associated with recruitment of DNMT3B to this locus (Marques et al. 2013). Comparison of the levels of TET2 and DNMT1 at different ER sites, for example through ChIP-seq, could provide insights into the potential balance between methylation and any demethylation occurring at these loci.

As part of future work, it may also be important to confirm which ER/TET2 sites are directly associated with TET2-regulated genes. This study has taken a proximity-based approach, assessing all TET2-regulated genes within a ± 50 kb window as potentially regulated by a particular ER/TET2 site. This is based on the finding that the majority of ER enhancers appear to lie within 10-100 kb of their target genes (Fullwood et al. 2009), however, many enhancers may be further away from their associated promoters. In addition, as described in Chapter 1, some ER binding events may not be functional, but may rather be the result of opportunistic ER binding (Carroll et al. 2006). Enhancers that are active in these cells

could be determined by performing ChIP-seq for active histone marks or monitoring the production of eRNAs. In conjunction with this, ER/TET2 sites could be linked to TET2-regulated genes through mapping chromatin loops before and after TET2 knockdown. These approaches might provide additional perspectives on the relationship between changes in DNA modifications and TET2-regulated expression of ER target genes. Furthermore, in addition to investigating coding genes, changes in the expression of non-coding RNAs could also be examined as endpoints of TET2 activity.

Adding further complexity to the potential relationship between TET2 and ER-mediated gene expression, as described in Chapter 1, TET2 may have additional functions related to transcription. These might include mediating protein-protein interactions between transcriptional regulators, or stabilising histone modifying complexes at gene regulatory sites (Deplus et al. 2013; Zhang et al. 2015; Zhang et al. 2017). It is possible that these additional functions of TET2 may contribute to its role in gene regulation and fine-tune its effects at different loci. To distinguish the catalytic and non-catalytic functions of TET2, the activity of TET2 lacking dioxygenase functions could be examined in ER+ breast cancer cells. To complement the acute knockdown system used in this work, this could involve TET2 silencing followed by overexpression of a catalytically dead TET2 mutant to assess whether this rescues the effects of TET2 loss on ER-regulated gene expression and global ER binding. In addition, with work to develop *TET2* knockout MCF7 cells using CRISPR currently in progress, CRISPR-generated TET2 catalytic mutants could provide a stable system for comparison. Overall, the findings summarised in this chapter provide novel insights into the mechanisms of ER biology. Placing these studies into the context of wider developments in both basic and translational research reveals several routes to build upon the work outlined in this thesis, towards the ultimate aim of developing more effective treatments for ER+ breast cancer.

6.4 Conclusions

In summary, the key points revealed by this study are:

- TET2 is a GATA3-regulated member of the ER complex in MCF7 cells, and is also an ER target gene.

- TET2 binding events constitute a near-total subset of ER binding events in ER+ breast cancer cell lines and PDX models according to ChIP-seq, implying potential genomic co-operation between these two proteins.
- Knockdown of TET2 in ER+ breast cancer cells depletes global ER binding and is associated with reduced cell growth and disrupted expression of several ER target genes. This suggests that TET2, as an ER-regulated component of the ER complex, may help to sustain ER-regulated transcription in breast cancer.
- TET2 knockdown does not appear to result in changes to global DNA methylation in ER+ breast cancer cells. However, oxidation of methylated DNA to 5-hydroxymethylcytosine (5hmC) is significantly reduced after TET2 depletion. This indicates a key role for TET2 in the production and maintenance of 5hmC at ER sites, providing a potential mechanism for TET2-mediated regulation of ER target genes.

Chapter 7

References

Adomas, A. B, Grimm, S. A., Malone, C., Takaku, M., Sims, J. K., Wade, P. A., Koboldt, D. C., et al. 2014. “Breast Tumor Specific Mutation in GATA3 Affects Physiological Mechanisms Regulating Transcription Factor Turnover.” *BMC Cancer* 14 (1): 278. https://doi.org/10.1186/1471-2407-14-278.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., and Mason, C. E. 2012. “MethylKit: A Comprehensive R Package for the Analysis of Genome-Wide DNA Methylation Profiles.” *Genome Biology* 13 (10): R87. https://doi.org/10.1186/gb-2012-13-10-r87.

Ali, R. H., Rueda, O. M., Chin, S. F., Curtis, C., Dunning, M. J., Aparicio, S. A. J. R, and Caldas, C. 2014. “Genome-Driven Integrated Classification of Breast Cancer Validated in over 7,500 Samples.” *Genome Biology* 15 (8): 431. https://doi.org/10.1186/s13059-014-0431-1.

Anolik, J. H., Klinge, C. M., Hilf, R., and Bambara, R. A. 1995. “Cooperative Binding of Estrogen Receptor to DNA Depends on Spacing of Binding Sites, Flanking Sequence, and Ligand.” *Biochemistry* 34 (8): 2511–20. https://doi.org/10.1021/bi00008a015.

Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X. Y., Sauter, G., Kallioniemi, O. P., Trent, J. M., and Meltzer, P. S.. 1997. “AIB1, a Steroid Receptor Coactivator Amplified in Breast and Ovarian Cancer.” *Science* 277 (5328): 965–68. https://doi.org/10.1126/science.277.5328.965.

Ariazi, E. A., Taylor, J. C., Black, M. A., Nicolas, E., Slifker M. J., Azzam, D. J., and Boyd, J. 2017. “A New Role for ERα: Silencing via DNA Methylation of Basal, Stem Cell, and EMT Genes.” *Molecular Cancer Research* 15 (2): 152–64. https://doi.org/10.1158/1541-7786.MCR-16-0283.

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. 2014. “5-Hydroxymethylcytosine Is a Predominantly Stable DNA Modification.” *Nature Chemistry* 6 (12): 1049–55. https://doi.org/10.1038/nchem.2064.

Bailey, T. L., and Elkan, C.. 1994. “Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers.” *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 2: 28–36. http://www.ncbi.nlm.nih.gov/pubmed/7584402.

Bailey, T. L. 2011. “DREME: Motif Discovery in Transcription Factor ChIP-Seq Data.” *Bioinformatics* 27 (12): 1653–59. https://doi.org/10.1093/bioinformatics/btr261.

Baker, M. E. 2013. “What are the physiological estrogens?” *Steroids* 78 (3): 337–340. https://doi.org/10.1016/j.steroids.2012.12.011.

Beard, C., Li, E., and Jaenisch, R. 1995. “Loss of Methylation Activates Xist in Somatic but Not in Embryonic Cells.” *Genes & Development* 9 (19): 2325–34. https://doi.org/10.1101/gad.9.19.2325.

Beatson, G. T. 1896. “ON THE TREATMENT OF INOPERABLE CASES OF CARCINOMA OF THE MAMMA: SUGGESTIONS FOR A NEW METHOD OF TREATMENT, WITH ILLUSTRATIVE CASES.” *The Lancet* 148 (3803): 162–65. https://doi.org/10.1016/S0140-6736(01)72384-7.

Belandia, B., Orford, R. L., Hurst, H. C., and Parker, M. G. 2002. “Targeting of SWI/SNF Chromatin Remodelling Complexes to Estrogen-Responsive Genes.” *The EMBO Journal* 21 (15): 4094–4103. https://doi.org/10.1093/emboj/cdf412.

Bird, A. 2002. “DNA Methylation Patterns and Epigenetic Memory.” *Genes & Development* 16 (1): 6–21. https://doi.org/10.1101/gad.947102.

Bird, A. P., and Wolffe, A. P. 1999. “Methylation-Induced Repression--Belts, Braces, and Chromatin.” *Cell* 99 (5): 451–54. https://doi.org/10.1016/s0092-8674(00)81532-9.

Bird, A. P., and Wolffe, A. P. 1999. “Methylation-Induced Repression— Belts, Braces, and Chromatin.” *Cell* 99 (5): 451–54. https://doi.org/10.1016/S0092-8674(00)81532-9.

Bonéy-Montoya, J., Ziegler, Y. S., Curtis, C. D., Montoya, J. A., and Nardulli, A. M. 2010. “Long-Range Transcriptional Control of Progesterone Receptor Gene Expression.” *Molecular Endocrinology* 24 (2): 346–58. https://doi.org/10.1210/me.2009-0429.

Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., Branco, M. R., Reik, R., and Balasubramanian, S. 2013. “Oxidative Bisulfite Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine.” *Nature Protocols* 8 (10): 1841–51. https://doi.org/10.1038/nprot.2013.115.

Bostick, M., Kim, J. K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. 2007. “UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells.” *Science* 317 (5845): 1760–64. https://doi.org/10.1126/science.1147939.

Bourdeau, V., Deschênes, J., Métivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J. H., and Mader, S. 2004. “Genome-Wide Identification of High-Affinity Estrogen Response Elements in Human and Mouse.” *Molecular Endocrinology* 18 (6): 1411–27. https://doi.org/10.1210/me.2003-0441.

Brisken, C., and O’Malley, B. 2010. “Hormone Action in the Mammary Gland.” *Cold Spring Harbor Perspectives in Biology* 2 (12): a003178–a003178. https://doi.org/10.1101/cshperspect.a003178.

Brisken, C., and Ataca, D. 2015. “Endocrine Hormones and Local Signals during the Development of the Mouse Mammary Gland.” *Wiley Interdisciplinary Reviews: Developmental Biology* 4 (3): 181–95. https://doi.org/10.1002/wdev.172.

Brophy, P. D., Rasmussen, M., Parida, M., Bonde, G., Darbro, B. W., Hong, X., Clarke, J. C., et al. 2017. “A Gene Implicated in Activation of Retinoic Acid Receptor Targets Is a Novel Renal Agenesis Gene in Humans.” *Genetics* 207 (1): 215–28. https://doi.org/10.1534/genetics.117.1125.

Bruna, A., Rueda, O. M., Greenwood, W., Serra, V., Garnett, M. J., and Caldas, C. 2016. “A Biobank of Breast Cancer Explants with Preserved Intra-Tumor Heterogeneity to Screen Anticancer Compounds In Brief.” *Cell* 167: 260–74. https://doi.org/10.1016/j.cell.2016.08.041.

Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engström, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. 1997. “Molecular Basis of Agonism and Antagonism in the Oestrogen Receptor.” *Nature* 389 (6652): 753–58. https://doi.org/10.1038/39645.

Cancer Genome Atlas Network. 2012. “Comprehensive Molecular Portraits of Human Breast Tumours.” *Nature* 490 (7418): 61–70. https://doi.org/10.1038/nature11412.

Cantor, A. B., and Orkin, S. H. 2005. “Coregulation of GATA Factors by the Friend of GATA (FOG) Family of Multitype Zinc Finger Proteins.” *Seminars in Cell & Developmental Biology* 16 (1): 117–28. https://doi.org/10.1016/j.semcdb.2004.10.006.

Carroll, J. S. 2016. “Mechanisms of Oestrogen Receptor (ER) Gene Regulation in Breast Cancer.” *European Journal of Endocrinology* 175 (1): R41–49. https://doi.org/10.1530/EJE-16-0124.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., et al. 2005. “Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1.” *Cell* 122 (1): 33–43. https://doi.org/10.1016/j.cell.2005.05.008.

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., et al. 2006. “Genome-Wide Analysis of Estrogen Receptor Binding Sites.” *Nature Genetics* 38 (11): 1289–97. https://doi.org/10.1038/ng1901.

Castet, A., Boulahtouf, A., Versini, G., Bonnet, S., Augereau, P., Vignon, F., Khochbin, S., Jalaguier, S., and Cavaillès, V. 2004. “Multiple Domains of the Receptor-Interacting Protein 140 Contribute to Transcription Inhibition.” *Nucleic Acids Research* 32 (6): 1957. https://doi.org/10.1093/NAR/GKH524.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., et al. 2012. “The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data.” Cancer Discovery 2 (5): 401–4. https://doi.org/10.1158/2159-8290.CD-12-0095.

Chen, D., Huang, S., and Stallcup, M. R. 2000. “Synergistic, P160 Coactivator-Dependent Enhancement of Estrogen Receptor Function by CARM1 and P300.” *Journal of Biological Chemistry* 275 (52): 40810–16. https://doi.org/10.1074/jbc.M005459200.

Chen, L., Lin, H., Zhou, W., He, C., Zhang, Z., Cheng, Z., Song, J., et al. 2018. “SNIP1 Recruits TET2 to Regulate C-MYC Target Genes and Cellular DNA Damage Response.” *Cell Reports* 25 (6): 1485-1500.e4. https://doi.org/10.1016/j.celrep.2018.10.028.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. 2013. “TET2 Promotes Histone O-GlcNAcylation during Gene Transcription.” *Nature* 493 (7433): 561–64. https://doi.org/10.1038/nature11742.

Chou, J., Provot, S., and Werb, Z. 2010. “GATA3 in Development and Cancer Differentiation: Cells GATA Have It!” *Journal of Cellular Physiology* 222 (1): 42–49. https://doi.org/10.1002/jcp.21943.

Ciesielski, P., Jóźwiak, P., Wójcik-Krowiranda, K., Forma, E., Cwonda, L., Szczepaniec, S., Bieńkiewicz, A., Bryś, M., and Krześlak, A. 2017. “Differential Expression of Ten-Eleven Translocation Genes in Endometrial Cancers.” *Tumor Biology* 39 (3): 101042831769501. https://doi.org/10.1177/1010428317695017.

Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., Clark, K. L., Burley, S. K., and Zaret, K. S. 1998. “Binding of the Winged-Helix Transcription Factor HNF3 to a Linker Histone Site on the Nucleosome.” *The EMBO Journal* 17 (1): 244–54. https://doi.org/10.1093/emboj/17.1.244.

Cirillo, L. A., Lin, F. L., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. S. 2002. “Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4.” *Molecular Cell* 9 (2): 279–89. https://doi.org/10.1016/S1097-2765(02)00459-8.

Clarke, R., Leonessa, F., Welch, J. N., and Skaar, T. C. 2001. “Cellular and Molecular Pharmacology of Antiestrogen Action and Resistance.” *Pharmacological Reviews* 53 (1): 25–71. http://www.ncbi.nlm.nih.gov/pubmed/11171938.

Cole, M. P., Jones, C. T., and Todd, I. D. 1971. “A New Anti-Oestrogenic Agent in Late Breast Cancer. An Early Clinical Appraisal of ICI46474.” *British Journal of Cancer* 25 (2): 270–75. https://doi.org/10.1038/bjc.1971.33.

Cole, P. A., and Robinson, C. H. 1990. “Mechanism and Inhibition of Cytochrome P-450 Aromatase.” *Journal of Medicinal Chemistry* 33 (11): 2933–42. https://doi.org/10.1021/jm00173a001.

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., et al. 2012. “The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups.” *Nature* 486 (7403): 346. https://doi.org/10.1038/nature10983.

Dauvois, S., White, R., and Parker, M. G. 1993. “The Antiestrogen ICI 182780 Disrupts Estrogen Receptor Nucleocytoplasmic Shuttling.” *Journal of Cell Science* 106 (Pt 4 (December): 1377–88. http://www.ncbi.nlm.nih.gov/pubmed/8126115.

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y., Markoulaki, S., Cheng, A. W., Gao, Q., et al. 2011. “Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development.” *Cell Stem Cell* 9 (2): 166–75. https://doi.org/10.1016/j.stem.2011.07.010.

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Inmaculada-Barrasa, M., Cheng, A. W., Gao, Q., et al. 2013. “Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development.” *Developmental Cell* 24 (3): 310–23. https://doi.org/10.1016/j.devcel.2012.12.015.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. 2002. “Capturing Chromosome Conformation.” *Science* 295 (5558): 1306–11. https://doi.org/10.1126/science.1067799.

Delage-Mourroux, R., Martini, P. G. V., Choi, I., Kraichely, D. M., Hoeksema, J., and Katzenellenbogen, B. S. 2000. “Analysis of Estrogen Receptor Interaction with a Repressor of Estrogen Receptor Activity (REA) and the Regulation of Estrogen Receptor Transcriptional Activity by REA.” *Journal of Biological Chemistry* 275 (46): 35848–56. https://doi.org/10.1074/jbc.M001327200.

Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Méndez, J., Murphy, N., Dawson, M. A., et al. 2013. “TET2 and TET3 Regulate GlcNAcylation and H3K4 Methylation through OGT and SET1/COMPASS.” *The EMBO Journal* 32 (5): 645–55. https://doi.org/10.1038/emboj.2012.357.

Deschênes, J., Bourdeau, V., White, J. H., and Mader, S. 2007. “Regulation of GREB1 Transcription by Estrogen Receptor Alpha through a Multipartite Enhancer Spread over 20 Kb of Upstream Flanking Sequences.” *The Journal of Biological Chemistry* 282 (24): 17335–39. https://doi.org/10.1074/jbc.C700030200.

DiRenzo, J., Shang, Y., Phelan, M., Sif, S., Myers, M., Kingston, R., and Brown, M. 2000. “BRG-1 Is Recruited to Estrogen-Responsive Promoters and Cooperates with Factors Involved in Histone Acetylation.” *Molecular and Cellular Biology* 20 (20): 7541–49. https://doi.org/10.1128/mcb.20.20.7541-7549.2000.

Dobin, A., and Gingeras, T. R. 2015. “Mapping RNA-Seq Reads with STAR.” *Current Protocols in Bioinformatics* 51 (September): 11.14.1-19. https://doi.org/10.1002/0471250953.bi1114s51.

Du, F., Yuan, P., Wang, T., Zhao, J., Zhao, Z., Luo, Y., and Xu, B. 2015. “The Significance and Therapeutic Potential of GATA3 Expression and Mutation in Breast Cancer: A Systematic Review.” *Medicinal Research Reviews* 35 (6): 1300–1315. https://doi.org/10.1002/med.21362.

Dutertre, M., and Smith, C. L. 2000. “Molecular Mechanisms of Selective Estrogen Receptor Modulator (SERM) Action.” *The Journal of Pharmacology and Experimental Therapeutics* 295 (2): 431–37. http://www.ncbi.nlm.nih.gov/pubmed/11046073.

Early Breast Cancer Trialists’ Collaborative Group (EBCTCG), Davies, C., Godwin, J., Gray, R., Clarke, M., Cutter, D., Darby, S., et al. 2011. “Relevance of Breast Cancer Hormone Receptors and Other Factors to the Efficacy of Adjuvant Tamoxifen: Patient-Level Meta-Analysis of Randomised Trials.” *Lancet (London, England)* 378 (9793): 771–84. https://doi.org/10.1016/S0140-6736(11)60993-8.

Eeckhoute, J., Krasnickas-Keeton, E., Lupien, M., Krum, S. A., Carroll, J. S., and Brown, M. 2007. “Positive Cross-Regulatory Loop Ties GATA-3 to Estrogen Receptor Alpha Expression in Breast Cancer.” *Cancer Research* 67 (13): 6477–83. https://doi.org/10.1158/0008-5472.CAN-07-0746.

Ellis, M. J., Suman, V. J., Hoog, J., Lin, L., Snider, J., Prat, A., Parker, J. S., et al. 2011. “Randomized Phase II Neoadjuvant Comparison Between Letrozole, Anastrozole, and Exemestane for Postmenopausal Women With Estrogen Receptor–Rich Stage 2 to 3 Breast Cancer: Clinical and Biomarker Outcomes and Predictive Value of the Baseline PAM50-Based Intrinsic Subtype—ACOSOG Z1031.” *Journal of Clinical Oncology* 29 (17): 2342–49. https://doi.org/10.1200/JCO.2010.31.6950.

Emmanuel, N., Lofgren, K. A., Peterson, E. A., Meier, D. R., Jung, E. H., and Kenny, P. A. 2018. “Mutant GATA3 Actively Promotes the Growth of Normal and Malignant Mammary Cells.” *Anticancer Research* 38 (8): 4435. https://doi.org/10.21873/ANTICANRES.12745.

ENCODE Project Consortium 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489 (7414): 57. https://doi.org/10.1038/NATURE11247.

Fackler, M. J., Umbricht, C. B., Williams, D., Argani, P., Cruz, L. A., Merino, V. F., Teo, W. W., et al. 2011. “Genome-Wide Methylation Analysis Identifies Genes Specific to Breast Cancer Hormone Receptor Status and Risk of Recurrence.” *Cancer Research* 71 (19): 6195–6207. https://doi.org/10.1158/0008-5472.CAN-11-1630.

Fan, M., Yan, P. S., Hartman-Frey, C., Chen, L., Paik, H., Oyer, S. L., Salisbury, J. D., et al. 2006. “Diverse Gene Expression and DNA Methylation Profiles Correlate with Differential Adaptation of Breast Cancer Cells to the Antiestrogens Tamoxifen and Fulvestrant.” *Cancer Research* 66 (24): 11954–66. https://doi.org/10.1158/0008-5472.CAN-06-1666.

Fawell, S. E., White, R., Hoare, S., Sydenham, M., Page, M., and Parker, M. G. 1990. “Inhibition of Estrogen Receptor-DNA Binding by the Pure Antiestrogen ICI 164,384 Appears to Be Mediated by Impaired Receptor Dimerization.” *Proceedings of the National Academy of Sciences* 87 (17): 6883–87. https://doi.org/10.1073/pnas.87.17.6883.

Fei, T., Peng, J., Xiao, T., Chen, C. H., Wu, A., H., J., Zang, C., Liu, S., and Brown, M. 2019. “Deciphering essential cistromes using genome-wide CRISPR screens” *PNAS* 116 (50): 25186-25195. https://doi.org/10.1073/pnas.1908155116.

Feng, Y., Manka, D., Wagner, K. U., and Khan, S. A. 2007. “Estrogen Receptor-Alpha Expression in the Mammary Epithelium Is Required for Ductal and Alveolar Morphogenesis in Mice.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (37): 14718–23. https://doi.org/10.1073/pnas.0706933104.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J. Z., Innocenti, C., Ferrier, P., et al. 2012. “CpG Islands and GC Content Dictate Nucleosome Depletion in a Transcription-Independent Manner at Mammalian Promoters.” *Genome Research* 22 (12): 2399–2408. https://doi.org/10.1101/gr.138776.112.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. 2015. “Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012.” *International Journal of Cancer* 136 (5): E359–86. https://doi.org/10.1002/ijc.29210.

Ficz, G., and Gribben, J. G. 2014. “Loss of 5-Hydroxymethylcytosine in Cancer: Cause or Consequence?” *Genomics* 104 (5): 352–57. https://doi.org/10.1016/J.YGENO.2014.08.017.

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P. A. 2006. “A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription.” *Molecular and Cellular Biology* 26 (1): 169–81. https://doi.org/10.1128/MCB.26.1.169-181.2006.

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., and Leonhardt H. 2011. “Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain.” Edited by Shuang-yong Xu. *PLoS ONE* 6 (6): e21306. https://doi.org/10.1371/journal.pone.0021306.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov. Y. L., et al. 2009. “An Oestrogen-Receptor-α-Bound Human Chromatin Interactome.” *Nature* 462 (7269): 58–64. https://doi.org/10.1038/nature08497.

Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J. C., and Sladek, R. 2009. “TFCat: The Curated Catalog of Mouse and Human Transcription Factors.” *Genome Biology* 10 (3): R29. https://doi.org/10.1186/gb-2009-10-3-r29.

García-Pedrero, J. M., Kiskinis, E., Parker, M. G., and Belandia, B. 2006. “The SWI/SNF Chromatin Remodeling Subunit BAF57 Is a Critical Regulator of Estrogen Receptor Function in Breast Cancer Cells.” *Journal of Biological Chemistry* 281 (32): 22656–64. https://doi.org/10.1074/jbc.M602561200.

Gardner, E. E., Lok, B. H., Schneeberger, V. E., Desmeules, P., Miles, L. A., Arnold, P. K., Ni, A., et al. 2017. “Chemosensitive Relapse in Small Cell Lung Cancer Proceeds through an EZH2-SLFN11 Axis.” *Cancer Cell* 31 (2): 286–99. https://doi.org/10.1016/j.ccell.2017.01.006.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. 2010. “Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates.” *PloS One* 5 (12): e15367. https://doi.org/10.1371/journal.pone.0015367.

Glont, S. E., Papachristou, E. K., Sawle, A., Holmes, K. A., Carroll, J. S., and Siersbaek, R. 2019. “Identification of ChIP-Seq and RIME Grade Antibodies for Estrogen Receptor Alpha.” Edited by Alessandro Weisz. *PLOS ONE* 14 (4): e0215340. https://doi.org/10.1371/journal.pone.0215340.

Glont, S. E., Chernukhin. I., and Carroll, J. S. 2019. “Comprehensive Genomic Analysis Reveals That the Pioneering Function of FOXA1 Is Independent of Hormonal Signaling.” *Cell Reports* 26. https://doi.org/10.1016/j.celrep.2019.02.036.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., et al. 2013. “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal.” Science Signaling 6 (269): pl1. https://doi.org/10.1126/scisignal.2004088.

Grant, C. E., Bailey, T. L., and Noble, W. S. 2011. “FIMO: Scanning for Occurrences of a given Motif.” *Bioinformatics (Oxford, England)* 27 (7): 1017–18. https://doi.org/10.1093/bioinformatics/btr064.

Gu, T., Guo F., Yang, H., Wu, H., Xu, G., Liu, W., Xie, Z., et al. 2011. “The Role of Tet3 DNA Dioxygenase in Epigenetic Reprogramming by Oocytes.” *Nature* 477 (7366): 606–10. https://doi.org/10.1038/nature10443.

Guan, W., Guyot, R., Samarut, J., Flamant, F., Wong, J., and Gauthier, K. C. 2017. “Methylcytosine Dioxygenase TET3 Interacts with Thyroid Hormone Nuclear Receptors and Stabilizes Their Association to Chromatin.” *Proceedings of the National Academy of Sciences* 114 (31): 8229–34. https://doi.org/10.1073/pnas.1702192114.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. 2007. “Quantifying Similarity between Motifs.” *Genome Biology* 8 (2): R24. https://doi.org/10.1186/gb-2007-8-2-r24.

Gustin, J. P., Miller, J., Farag, M., Rosen, D. M., Thomas, M., Scharpf, R. B., and Lauring, J. 2017. “GATA3 Frameshift Mutation Promotes Tumor Growth in Human Luminal Breast Cancer Cells and Induces Transcriptional Changes Seen in Primary GATA3 Mutant Breast Cancers.” *Oncotarget* 8 (61): 103415–27. https://doi.org/10.18632/oncotarget.21910.

Györffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. 2010. “An Online Survival Analysis Tool to Rapidly Assess the Effect of 22,277 Genes on Breast Cancer Prognosis Using Microarray Data of 1,809 Patients.” *Breast Cancer Research and Treatment* 123 (3): 725–31. https://doi.org/10.1007/s10549-009-0674-9.

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T., and Surani, M. A. 2013. “Germline DNA Demethylation Dynamics and Imprint Erasure through 5-hyroxymethylcytosine.” *Science* 339 (6118): 448–52. https://doi.org/10.1126/science.1229277

Hah, N., Murakami, S., Nagari, A., Danko, C. G., and Kraus, W. L. 2013. “Enhancer Transcripts Mark Active Estrogen Receptor Binding Sites.” *Genome Research* 23 (8): 1210–23. https://doi.org/10.1101/gr.152306.112.

Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., et al. 2013. “Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis.” *Cell Reports* 3 (2): 291–300. https://doi.org/10.1016/j.celrep.2013.01.011.

Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., and Surani, M. A. 2010. “Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway.” *Science (New York, N.Y.)* 329 (5987): 78–82. https://doi.org/10.1126/science.1187945.

Hardivillé, S., and Hart, G. W. 2014. “Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation.” *Cell Metabolism* 20 (2): 208–13. https://doi.org/10.1016/J.CMET.2014.07.014.

Harrod, A., Fulton, J., Nguyen, V. T. M., Periyasamy, M., Ramos-Garcia, L., Lai, C., Metodieva, G., et al. 2017. “Genomic Modelling of the ESR1 Y537S Mutation for Evaluating Function and New Therapeutic Approaches for Metastatic Breast Cancer.” *Oncogene* 36 (16): 2286–96. https://doi.org/10.1038/onc.2016.382.

Hashimoto, H., Liu, , Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., and Cheng, X. 2012. “Recognition and Potential Mechanisms for Replication and Erasure of Cytosine Hydroxymethylation.” *Nucleic Acids Research* 40 (11): 4841–49. https://doi.org/10.1093/nar/gks155.

Haskell, S. G. 2003. “Selective Estrogen Receptor Modulators.” *Southern Medical Journal* 96 (5): 469–76. https://doi.org/10.1097/01.SMJ.0000051146.93190.4A.

Hassan, H. M., Kolendowski, B., Isovic, M., Bose, K., Dranse, H. J., Sampaio, A. V., Underhill, T. M., and Torchia, J. 2017. “Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex.” *Cell Reports* 19 (8): 1685–97. https://doi.org/10.1016/j.celrep.2017.05.007.

He, Y., Li, B., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., et al. 2011. “Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA.” *Science (New York, N.Y.)* 333 (6047): 1303–7. https://doi.org/10.1126/science.1210944.

Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G. 1997. “A Signature Motif in Transcriptional Co-Activators Mediates Binding to Nuclear Receptors.” *Nature* 387 (6634): 733–36. https://doi.org/10.1038/42750.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., et al. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459 (7243): 108–12. https://doi.org/10.1038/nature07829.

Heinz, S., Romanoski, C. E., Benner, C., and Glass, C. K. 2015. “The Selection and Function of Cell Type-Specific Enhancers.” *Nature Reviews Molecular Cell Biology* 16 (3): 144–54. https://doi.org/10.1038/nrm3949.

Herlin, M. K., Le, V. Q., Højland, A. T., Ernst, A., Okkels, H., Petersen, A. C., Petersen, M. B., and Pedersen, I. S. 2019. “Whole-Exome Sequencing Identifies a GREB1L Variant in a Three-Generation Family with Müllerian and Renal Agenesis: A Novel Candidate Gene in Mayer–Rokitansky–Küster–Hauser (MRKH) Syndrome. A Case Report.” *Human Reproduction* 34 (9): 1838–46. https://doi.org/10.1093/humrep/dez126.

Hermann, A., Goyal, R., and Jeltsch A. 2004. “The Dnmt1 DNA-(Cytosine-C5)-Methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites.” *Journal of Biological Chemistry* 279 (46): 48350–59. https://doi.org/10.1074/jbc.M403427200.

Hernandez-Blazquez, F. J., Habib, M., Dumollard, J. M., Barthelemy, C., Benchaib, M., de Capoa, A., and Niveleau, A. 2000. “Evaluation of Global DNA Hypomethylation in Human Colon Cancer Tissues by Immunohistochemistry and Image Analysis.” *Gut* 47 (5): 689–93. https://doi.org/10.1136/gut.47.5.689.

Hill, V. K., Ricketts, C., Bieche, I., Vacher, S., Gentle, D., Lewis, C., Maher, E. R., and Latif, F. 2011. “Genome-Wide DNA Methylation Profiling of CpG Islands in Breast Cancer Identifies Novel Genes Associated with Tumorigenicity.” *Cancer Research* 71 (8): 2988–99. https://doi.org/10.1158/0008-5472.CAN-10-4026.

Holmes, K. A., Hurtado, A., Brown, G. D., Launchbury, R., Ross-Innes, C. S., Hadfield, J., Odom, D. T., and Carroll, J. S. 2012. “Transducin-like Enhancer Protein 1 Mediates Estrogen Receptor Binding and Transcriptional Activity in Breast Cancer Cells.” *Proceedings of the National Academy of Sciences* 109 (8): 2748–53. https://doi.org/10.1073/pnas.1018863108.

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., and Ren, B. 2013. “Epigenetic Memory at Embryonic Enhancers Identified in DNA Methylation Maps from Adult Mouse Tissues.” *Nature Genetics* 45 (10): 1198–1206. https://doi.org/10.1038/ng.2746.

Hon, G. C., Song, C. X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., Yen, C. A., et al. 2014. “5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation.” *Molecular Cell* 56 (2): 286–97. https://doi.org/10.1016/j.molcel.2014.08.026.

Hong, H., Kohli, K., Garabedian, M. J., and Stallcup, M, R. 1997. “GRIP1, a Transcriptional Coactivator for the AF-2 Transactivation Domain of Steroid, Thyroid, Retinoid, and Vitamin D Receptors.” *Molecular and Cellular Biology* 17 (5): 2735–44. https://doi.org/10.1128/mcb.17.5.2735.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., and Xu, Y. 2013. “Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation.” *Cell* 155 (7): 1545–55. https://doi.org/10.1016/J.CELL.2013.11.020.

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., et al. 2015. “Structural Insight into Substrate Preference for TET-Mediated Oxidation.” *Nature* 527 (7576): 118–22. https://doi.org/10.1038/nature15713.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. 2009. “Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources.” *Nature Protocols* 4 (1): 44–57. https://doi.org/10.1038/nprot.2008.211.

Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H. C., and Lempicki, R. A. 2007. “The DAVID Gene Functional Classification Tool: A Novel Biological Module-Centric Algorithm to Functionally Analyze Large Gene Lists.” *Genome Biology* 8 (9): R183. https://doi.org/10.1186/gb-2007-8-9-r183.

Huang, P., Chandra, V., and Rastinejad, F. 2010. “Structural Overview of the Nuclear Receptor Superfamily: Insights into Physiology and Therapeutics.” *Annual Review of Physiology* 72: 247–72. https://doi.org/10.1146/annurev-physiol-021909-135917.

Huang, Y., Wang, G., Liang, Z., Yang, Y., Cui, L., and Liu, C. Y. 2016. “Loss of Nuclear Localization of TET2 in Colorectal Cancer.” *Clinical Epigenetics* 8 (1): 9. https://doi.org/10.1186/s13148-016-0176-7.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D., and Carroll, J. S. 2011. “FOXA1 Is a Key Determinant of Estrogen Receptor Function and Endocrine Response.” *Nature Genetics* 43 (1): 27–33. https://doi.org/10.1038/ng.730.

Inoue, A., and Zhang, Y. 2011. “Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos.” *Science* 334 (6053): 194–194. https://doi.org/10.1126/science.1212483.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. 2011. “Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine.” *Science (New York, N.Y.)* 333 (6047): 1300–1303. https://doi.org/10.1126/science.1210597.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., and Reik, W. 2013. “A Screen for Hydroxymethylcytosine and Formylcytosine Binding Proteins Suggests Functions in Transcription and Chromatin Regulation.” *Genome Biology* 14 (10): R119. https://doi.org/10.1186/gb-2013-14-10-r119.

Iurlaro, M., McInroy, G. R., Burgess, H. E., Dean, W., Raiber, E. A., Bachman, M., Beraldi, D., Balasubramanian, S., and Reik, W. 2016. “In Vivo Genome-Wide Profiling Reveals a Tissue-Specific Role for 5-Formylcytosine.” *Genome Biology* 17 (1): 141. https://doi.org/10.1186/s13059-016-1001-5.

Iyer, L. M., Zhang, D., Burroughs, A. M., and Aravind, L. 2013. “Computational Identification of Novel Biochemical Systems Involved in Oxidation, Glycosylation and Other Complex Modifications of Bases in DNA.” *Nucleic Acids Research* 41 (16): 7635–55. https://doi.org/10.1093/nar/gkt573.

Jaenisch, R., Schnieke, A., and Harbers, K. 1985. “Treatment of Mice with 5-Azacytidine Efficiently Activates Silent Retroviral Genomes in Different Tissues.” *Proceedings of the National Academy of Sciences* 82 (5): 1451–55. https://doi.org/10.1073/pnas.82.5.1451.

Jakacka, M., Ito, M., Weiss, J., Chien, P. Y., Gehm, B. D., and Jameson, J. L. 2001. “Estrogen Receptor Binding to DNA Is Not Required for Its Activity through the Nonclassical AP1 Pathway.” *Journal of Biological Chemistry* 276 (17): 13615–21. https://doi.org/10.1074/jbc.M008384200.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G. A., Rodriguez, B., Zhang, X., et al. 2014. “Large Conserved Domains of Low DNA Methylation Maintained by Dnmt3a.” *Nature Genetics* 46 (1): 17–23. https://doi.org/10.1038/ng.2836.

Ji, D., Lin, K., Song, J., and Wang, Y. 2014. “Effects of Tet-Induced Oxidation Products of 5-Methylcytosine on Dnmt1- and DNMT3a-Mediated Cytosine Methylation.” *Molecular BioSystems* 10 (7): 1749. https://doi.org/10.1039/c4mb00150h.

Jin, S. G, Zhang, Z. M., Dunwell, T. L., Xu, G. L., Song, J., Pfeifer, G. P., and Et, J. 2016. “Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration Accession Numbers GSE56821 5EXH.” *CellReports* 14: 493–505. https://doi.org/10.1016/j.celrep.2015.12.044.

Johnson, K. C., Houseman, E. A., King, J. E., von Herrmann, K. M., Fadul, C. E., and Christensen, B. C. 2016. “5-Hydroxymethylcytosine Localizes to Enhancer Elements and Is Associated with Survival in Glioblastoma Patients.” *Nature Communications* 7: 13177. https://doi.org/10.1038/ncomms13177.

Jones, P. L., Jan Veenstra, G. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. 1998. “Methylated DNA and MeCP2 Recruit Histone Deacetylase to Repress Transcription.” *Nature Genetics* 19 (2): 187–91. https://doi.org/10.1038/561.

Jordan, V. C. 2003. “Tamoxifen: A Most Unlikely Pioneering Medicine.” *Nature Reviews Drug Discovery* 2 (3): 205–13. https://doi.org/10.1038/nrd1031.

Joseph, R., Orlov. Y. L., Huss, M., Sun, W., Kong, S. L., Ukil, L., Pan, Y. F., et al. 2010. “Integrative Model of Genomic Factors for Determining Binding Site Selection by Estrogen Receptor-α.” *Molecular Systems Biology* 6 (1): 456. https://doi.org/10.1038/msb.2010.109.

Jozwik, K. M., and Carroll, J. S. 2012. “Pioneer Factors in Hormone-Dependent Cancers.” *Nature Reviews. Cancer* 12 (6): 381–85. https://doi.org/10.1038/nrc3263.

Kadauke, S., and Blobel, G. A. 2009. “Chromatin Loops in Gene Regulation.” *Biochimica et Biophysica Acta* 1789 (1): 17–25. https://doi.org/10.1016/j.bbagrm.2008.07.002.

Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., Benes, V., Gannon, F., and Reid, G. 2008. “Transient Cyclical Methylation of Promoter DNA.” *Nature* 452 (7183): 112–15. https://doi.org/10.1038/nature06640.

Karnuta, J. M., and Scacheri, P. C. 2018. “Enhancers: Bridging the Gap between Gene Control and Human Disease.” *Human Molecular Genetics* 27 (R2): R219–27. https://doi.org/10.1093/hmg/ddy167.

Kellinger, M. W., Song, C. X., Chong, J., Lu, X. Y., He, C., and Wang, D. 2012. “5-Formylcytosine and 5-Carboxylcytosine Reduce the Rate and Substrate Specificity of RNA Polymerase II Transcription.” *Nature Structural & Molecular Biology* 19 (8): 831–33. https://doi.org/10.1038/nsmb.2346.

Key, T. J., Verkasalo, P. K., and Banks, E. 2001. “Epidemiology of Breast Cancer.” *The Lancet. Oncology* 2 (3): 133–40. https://doi.org/10.1016/S1470-2045(00)00254-0.

Klinge, C. M. 2001. “Estrogen Receptor Interaction with Estrogen Response Elements.” *Nucleic Acids Research* 29 (14): 2905–19. https://doi.org/10.1093/nar/29.14.2905.

Klose, R. J., and Bird, A. P. 2006. “Genomic DNA Methylation: The Mark and Its Mediators.” *Trends in Biochemical Sciences* 31 (2): 89–97. https://doi.org/10.1016/J.TIBS.2005.12.008.

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., et al. 2013. “Modulation of TET2 Expression and 5-Methylcytosine Oxidation by the CXXC Domain Protein IDAX.” *Nature* 497 (7447): 122–26. https://doi.org/10.1038/nature12052.

Koh, S. S., Chen, D., Lee, Y. H., and Stallcup, M. R. 2001. “Synergistic Enhancement of Nuclear Receptor Function by P160 Coactivators and Two Coactivators with Protein Methyltransferase Activities.” *The Journal of Biological Chemistry* 276 (2): 1089–98. https://doi.org/10.1074/jbc.M004228200.

Kong, S. L., Li, G., Loh, S. L., Sung, W. K., and Liu, E. T. 2014. “Cellular Reprogramming by the Conjoint Action of ER , FOXA1, and GATA3 to a Ligand-Inducible Growth State.” *Molecular Systems Biology* 7 (1): 526–526. https://doi.org/10.1038/msb.2011.59.

Korde, L. A., Zujewski, J. A., Kamin, L., Giordano, S., Domchek, S., Anderson, W. F., Bartlett, J. M. S., et al. 2010. “Multidisciplinary Meeting on Male Breast Cancer: Summary and Research Recommendations.” *Journal of Clinical Oncology* 28 (12): 2114–22. https://doi.org/10.1200/JCO.2009.25.5729.

Kosmider, O., Gelsi-Boyer, V., Ciudad, M., Racoeur, C., Jooste, V., Vey, N., Quesnel, B., et al. 2009. “TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia.” Haematologica 94 (12): 1676-81. https://doi.org/10.3324/haematol.2009.011205.

Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D., and Werb, Z. 2006. “GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland.” *Cell* 127 (5): 1041–55. https://doi.org/10.1016/j.cell.2006.09.048.

Krueger, F., and Andrews, S. R. 2011. “Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications.” *Bioinformatics* 27 (11): 1571–72. https://doi.org/10.1093/bioinformatics/btr167.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H., and Koike, K. 2012. “Loss of 5-Hydroxymethylcytosine Is Accompanied with Malignant Cellular Transformation.” *Cancer Science* 103 (4): 670–76. https://doi.org/10.1111/j.1349-7006.2012.02213.x.

Kuiper, G. G., Enmark, E., Pelto-Huikko, M., Nilsson, S., and Gustafsson, J. A. 1996. “Cloning of a Novel Receptor Expressed in Rat Prostate and Ovary.” *Proceedings of the National Academy of Sciences* 93 (12): 5925–30. https://doi.org/10.1073/pnas.93.12.5925.

Kulis, M., and Esteller, M.. 2010. “DNA Methylation and Cancer.” In *Advances in Genetics*, 70:27–56. https://doi.org/10.1016/B978-0-12-380866-0.60002-2.

Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R., and Chambon, P. 1987. “Functional Domains of the Human Estrogen Receptor.” *Cell* 51 (6): 941–51. https://doi.org/10.1016/0092-8674(87)90581-2.

Kushner, P. J., Agard, D. A., Greene, G. L., Scanlan, T. S., Shiau, A. K., Uht, R. M., and Webb, P. 2000. “Estrogen Receptor Pathways to AP-1.” *The Journal of Steroid Biochemistry and Molecular Biology* 74 (5): 311–17. http://www.ncbi.nlm.nih.gov/pubmed/11162939.

Lacroix, M., and Leclercq, G. 2004. “About GATA3, HNF3A, and XBP1, Three Genes Co-Expressed with the Oestrogen Receptor-Alpha Gene (ESR1) in Breast Cancer.” *Molecular and Cellular Endocrinology* 219 (1–2): 1–7. https://doi.org/10.1016/j.mce.2004.02.021.

Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., and Giguere, V. 2005. “Location Analysis of Estrogen Receptor Target Promoters Reveals That FOXA1 Defines a Domain of the Estrogen Response.” *Proceedings of the National Academy of Sciences* 102 (33): 11651–56. https://doi.org/10.1073/pnas.0505575102.

Langmead, B., and Salzberg, S. L. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lazar, M. A. 2003. “Nuclear Receptor Corepressors.” *Nuclear Receptor Signaling* 1: 1. https://doi.org/10.1621/nrs.01001.

Leong, H., Sloan, J. R., Nash, P. D., and Greene, G. L. 2005. “Recruitment of Histone Deacetylase 4 to the N-Terminal Region of Estrogen Receptor α.” *Molecular Endocrinology* 19 (12): 2930–42. https://doi.org/10.1210/me.2005-0178.

Lercher, L., Raj, R., Patel, N. A., Price, J., Mohammed, S., Robinson, C. V., Schofield, C. J., and Davis, B. G. 2015. “Generation of a Synthetic GlcNAcylated Nucleosome Reveals Regulation of Stability by H2A-Thr101 GlcNAcylation.” *Nature Communications* 6. https://doi.org/10.1038/NCOMMS8978.

Leu, Y. W., Yan, P. S., Fan, M., Jin, V. X., Liu, J. C., Curran, E. M., Welshons, W. V., et al. 2004. “Loss of Estrogen Receptor Signaling Triggers Epigenetic Silencing of Downstream Targets in Breast Cancer.” *Cancer Research* 64 (22): 8184–92. https://doi.org/10.1158/0008-5472.CAN-04-2045.

Li, C. I. 2003. “Trends in Incidence Rates of Invasive Lobular and Ductal Breast Carcinoma.” *JAMA* 289 (11): 1421. https://doi.org/10.1001/jama.289.11.1421.

Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., Yang, F. C., and Xu, M. 2011. “Deletion of Tet2 in Mice Leads to Dysregulated Hematopoietic Stem Cells and Subsequent Development of Myeloid Malignancies.” Blood 118 (17): 4509–18. https://doi.org/10.1182/blood-2010-12-325241.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., et al. 2013. “Functional Roles of Enhancer RNAs for Oestrogen-Dependent Transcriptional Activation.” *Nature* 498 (7455): 516–20. https://doi.org/10.1038/nature12210.

Li, X., Yao, B., Chen, L., Kang, Y., Li, Y., Cheng, Y., Li, L., et al. 2017. “Ten-Eleven Translocation 2 Interacts with Forkhead Box O3 and Regulates Adult Neurogenesis.” *Nature Communications* 8 (June): 15903. https://doi.org/10.1038/ncomms15903.

Li, R., Zhou, Y., Cao, Z., Liu, L., Wang, J., Chen, Z., Xing, W., et al. 2018. “TET2 Loss Dysregulates the Behavior of Bone Marrow Mesenchymal Stromal Cells and Accelerates Tet2-/--Driven Myeloid Malignancy Progression.” Stem Cell Reports 10 (1): 166–79. https://doi.org/10.1016/j.stemcr.2017.11.019.

Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., et al. 2012. “Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma.” *Cell* 150 (6): 1135–46. https://doi.org/10.1016/j.cell.2012.07.033.

Liebler, D. C., and Zimmerman, L. J. 2013. “Targeted Quantitation of Proteins by Mass Spectrometry.” *Biochemistry* 52 (22): 3797–3806. https://doi.org/10.1021/bi400110b.

Lin, C. Y., Vega, V. B., Thomsen, J. S., Zhang, T., Kong, S. L., Xie, M., Chiu, K. P., et al. 2007. “Whole-Genome Cartography of Estrogen Receptor α Binding Sites.” *PLoS Genetics* 3 (6): e87. https://doi.org/10.1371/journal.pgen.0030087.

Lindahl, T., and Wood, R. D. 1999. “Quality Control by DNA Repair.” *Science* 286 (5446): 1897–1905. https://doi.org/10.1126/science.286.5446.1897.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., et al. 2009. “Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences.” *Nature* 462 (7271): 315–22. https://doi.org/10.1038/nature08514.

Liu, J., Prager–van der Smissen, W. J. C., Look, M. P., Sieuwerts, A. M., Smid, M., Meijer–van Gelder, M. E., Foekens, J. A., Hollestelle, A., and Martens, J. W. M. 2016. “GATA3 MRNA Expression, but Not Mutation, Associates with Longer Progression-Free Survival in ER-Positive Breast Cancer Patients Treated with First-Line Tamoxifen for Recurrent Disease.” *Cancer Letters* 376 (1): 104–9. https://doi.org/10.1016/j.canlet.2016.03.038.

Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M. J., Song, X., et al. 2014. “Enhancer Activation Requires Trans-Recruitment of a Mega Transcription Factor Complex.” *Cell* 159 (2): 358–73. https://doi.org/10.1016/J.CELL.2014.08.027.

Livak, K. J., and Schmittgen, T. D. 2001. “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method.” *Methods (San Diego, Calif.)* 25 (4): 402–8. https://doi.org/10.1006/meth.2001.1262.

López-Moyado, I. F., Tsagaratou, A., Yuita, H., Seo, H., Delatte, B., Heinz, S., Benner, C., and Rao, A. 2019. “Paradoxical association of TET loss of function with genome-wide DNA hypomethylation.” PNAS. 116 (34) 16933–16942. https://doi.org/10.1073/pnas.1903059116.

Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., and Downing, J. R. 2003. “TET1, a Member of a Novel Protein Family, Is Fused to MLL in Acute Myeloid Leukemia Containing the t(10;11)(Q22;Q23).” *Leukemia* 17 (3): 637–41. https://doi.org/10.1038/sj.leu.2402834.

Lou, H., Li, H., Ho, K. J., Cai, L. L., Huang, A. S., Shank, T. R., Verneris, M. R., Nickerson, M. L., Dean, M., and Anderson, S. K. 2019. “The Human TET2 Gene Contains Three Distinct Promoter Regions With Differing Tissue and Developmental Specificities.” *Frontiers in Cell and Developmental Biology* 7 (June): 99. https://doi.org/10.3389/fcell.2019.00099.

Love, M. I, Huber, W., and Anders, S. 2014a. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. 2014. “Role of Tet Proteins in Enhancer Activity and Telomere Elongation.” *Genes & Development* 28 (19): 2103–19. https://doi.org/10.1101/gad.248005.114.

Lubahn, D. B., Moyer, J. S., Golding, T. S., Couse, J. F., Korach, K. S., and Smithies, O. 1993. “Alteration of Reproductive Function but Not Prenatal Sexual Development after Insertional Disruption of the Mouse Estrogen Receptor Gene.” *Proceedings of the National Academy of Sciences* 90 (23): 11162–66. https://doi.org/10.1073/pnas.90.23.11162.

Luo, C., Hajkova, P., and Ecker, J. R. 2018. “Dynamic DNA Methylation: In the Right Place at the Right Time.” *Science* 361 (6409): 1336–40. https://doi.org/10.1126/science.aat6806.

Lupien, M., Eeckhoute, J., Meyer, C. A., Krum, S. A., Rhodes, D. R., Liu, X. S., and Brown, M. 2009. “Coactivator Function Defines the Active Estrogen Receptor Alpha Cistrome.” *Molecular and Cellular Biology* 29 (12): 3413–23. https://doi.org/10.1128/MCB.00020-09.

Ma, H., Baumann, C. T., Li, H., Strahl, B. D., Rice, R., Jelinek, M. A., Aswad, D. W., Allis, C. D., Hager, G. L., and Stallcup, M. R. 2001. “Hormone-Dependent, CARM1-Directed, Arginine-Specific Methylation of Histone H3 on a Steroid-Regulated Promoter.” *Current Biology : CB* 11 (24): 1981–85. https://doi.org/10.1016/s0960-9822(01)00600-5.

Ma, S., Wan, X., Deng, Z., Shi, L., Hao, C., Zhou, Z., Zhou, C., et al. 2017. “Epigenetic Regulator CXXC5 Recruits DNA Demethylase Tet2 to Regulate TLR7/9-Elicited IFN Response in PDCs.” *Journal of Experimental Medicine* 214 (5): 1471–91. https://doi.org/10.1084/JEM.20161149.

Madak-Erdogan, Z., Kieser, K. J, Kim, S. H., Komm, B., Katzenellenbogen, J. A., and Katzenellenbogen, B. S. 2008. “Nuclear and Extranuclear Pathway Inputs in the Regulation of Global Gene Expression by Estrogen Receptors.” Molecular Endocrinology 22 (9): 2116–27. https://doi.org/10.1210/me.2008-0059.

Madzo, J., Liu, H., Rodriguez, A., Vasanthakumar, A., Sundaravel, S., Caces, D. B. D., Looney, T. J, et al. 2014. “Hydroxymethylation at Gene Regulatory Regions Directs Stem/Early Progenitor Cell Commitment during Erythropoiesis.” *Cell Reports* 6 (1): 231–44. https://doi.org/10.1016/j.celrep.2013.11.044.

Magnani, L., Ballantyne, E. B., Zhang, X., and Lupien, M. 2011. “PBX1 Genomic Pioneer Function Drives ERα Signaling Underlying Progression in Breast Cancer.” Edited by Bruce E. Clurman. *PLoS Genetics* 7 (11): e1002368. https://doi.org/10.1371/journal.pgen.1002368.

Magnani, L., Patten, D. K., Nguyen, V. T. M., Hong, S. P., Steel, J. H., Patel, N., Lombardo, Y., et al. 2015. “The Pioneer Factor PBX1 Is a Novel Driver of Metastatic Progression in ER-Positive Breast Cancer.” *Oncotarget* 6 (26): 21878–91. https://doi.org/10.18632/oncotarget.4243.

Maiti, A., and Drohat, A. C. 2011. “Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine.” *Journal of Biological Chemistry* 286 (41): 35334–38. https://doi.org/10.1074/jbc.C111.284620.

Marchal, C., and Miotto, B. 2015. “Emerging Concept in DNA Methylation: Role of Transcription Factors in Shaping DNA Methylation Patterns.” *Journal of Cellular Physiology* 230 (4): 743–51. https://doi.org/10.1002/jcp.24836.

Marques, M., Laflamme, L., and Gaudreau, L. 2013. “Estrogen Receptor α Can Selectively Repress Dioxin Receptor-Mediated Gene Expression by Targeting DNA Methylation.” *Nucleic Acids Research* 41 (17): 8094–8106. https://doi.org/10.1093/nar/gkt595.

Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M. S., Ghosh, D., Chinnaiyan, A. M., and Kleer, C. G. 2005. “Identification of GATA3 as a Breast Cancer Prognostic Marker by Global Gene Expression Meta-Analysis.” *Cancer Research* 65 (24): 11259–64. https://doi.org/10.1158/0008-5472.CAN-05-2495.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., et al. 2008. “Cyclical DNA Methylation of a Transcriptionally Active Promoter.” *Nature* 452 (7183): 45–50. https://doi.org/10.1038/nature06544.

Miller, W. R. 2003. “Aromatase Inhibitors: Mechanism of Action and Role in the Treatment of Breast Cancer.” *Seminars in Oncology* 30: 3–11. https://doi.org/10.1016/S0093-7754(03)00302-6.

Mohammed, H., and Carroll, J. S. 2013. “Approaches for Assessing and Discovering Protein Interactions in Cancer.” *Molecular Cancer Research* 11 (11): 1295–1302. https://doi.org/10.1158/1541-7786.MCR-13-0454.

Mohammed, H., D’Santos, C., Serandour, A. A., Ali, H. R., Brown, G. D., Atkins, A., Rueda, O. M., et al. 2013. “Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor.” *Cell Reports* 3 (2): 342–49. https://doi.org/10.1016/j.celrep.2013.01.010.

Mohammed, H., Taylor, C., Brown, G. D., Papachristou, E. K., Carroll, J. S., and D’Santos, C. S. 2016. “Rapid Immunoprecipitation Mass Spectrometry of Endogenous Proteins (RIME) for Analysis of Chromatin Complexes.” *Nature Protocols* 11 (2): 316–26. https://doi.org/10.1038/nprot.2016.020.

Mohandas, T., Sparkes, R., and Shapiro, L. 1981. “Reactivation of an Inactive Human X Chromosome: Evidence for X Inactivation by DNA Methylation.” *Science* 211 (4480): 393–96. https://doi.org/10.1126/science.6164095.

Montagner, S., Leoni, C., Emming, S., Della Chiara, G., Balestrieri, C., Barozzi, I., Piccolo, V., et al. 2016. “TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-Catalytic Activities.” *Cell Reports* 15 (7): 1566–79. https://doi.org/10.1016/j.celrep.2016.04.044.

Montalbán-Loro, R., Lozano-Ureña, A., Ito, M., Krueger, C., Reik, W., Ferguson-Smith, A. C., and Ferrón, S. R. 2019. “TET3 Prevents Terminal Differentiation of Adult NSCs by a Non-Catalytic Action at Snrpn.” *Nature Communications* 10 (1): 1726. https://doi.org/10.1038/s41467-019-09665-1.

Moore, L. D, Le, T., and Fan, G. 2013. “DNA Methylation and Its Basic Function.” *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 38 (1): 23–38. https://doi.org/10.1038/npp.2012.112.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M. E., et al. 2011. “Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation.” *Cancer Cell* 20 (1): 11–24. https://doi.org/10.1016/j.ccr.2011.06.001.

Mourad, R., Hsu, P. Y., Juan, L., Shen, C., Koneru, P., Lin, H., Liu, Y., Nephew, K., Huang, T. H., and Li, L. 2014. “Estrogen Induces Global Reorganization of Chromatin Structure in Human Breast Cancer Cells.” *PloS One* 9 (12): e113354. https://doi.org/10.1371/journal.pone.0113354.

Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y., and Kohli, R. M. 2012. “AID/APOBEC Deaminases Disfavor Modified Cytosines Implicated in DNA Demethylation.” *Nature Chemical Biology* 8 (9): 751. https://doi.org/10.1038/NCHEMBIO.1042.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. 1998. “Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex.” *Nature* 393 (6683): 386–89. https://doi.org/10.1038/30764.

Neish, A., Anderson, S. F., Schlegel, B. P., Wei, W., and Parvin, J. D. 1998. “Factors Associated with the Mammalian RNA Polymerase II Holoenzyme.” *Nucleic Acids Research* 26 (3): 847–53. https://doi.org/10.1093/nar/26.3.847.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. 2013. “Genome-Wide Analysis Identifies a Functional Association of Tet1 and Polycomb Repressive Complex 2 in Mouse Embryonic Stem Cells.” *Genome Biology* 14 (8): R91. https://doi.org/10.1186/gb-2013-14-8-r91.

Onate, S. A., Tsai, S. Y., Tsai, M. J., and O’Malley, B. W. 1995. “Sequence and Characterization of a Coactivator for the Steroid Hormone Receptor Superfamily.” *Science* 270 (5240): 1354–57. https://doi.org/10.1126/science.270.5240.1354.

O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44 (D1): D733-45. https://doi.org/10.1093/nar/gkv1189.

Palaparti, A., Baratz, A., and Stifani, S. 1997. “The Groucho/Transducin-like Enhancer of Split Transcriptional Repressors Interact with the Genetically Defined Amino-Terminal Silencing Domain of Histone H3.” *Journal of Biological Chemistry* 272 (42): 26604–10. https://doi.org/10.1074/jbc.272.42.26604.

Palstra, R. J., and Grosveld, F. 2012. “Transcription Factor Binding at Enhancers: Shaping a Genomic Regulatory Landscape in Flux.” *Frontiers in Genetics* 3: 195. https://doi.org/10.3389/fgene.2012.00195.

Pan, Y. F, Wansa, K. D., Liu, M. H., Zhao, B., Hong, S. Z., Tan, P. Y., Lim, K. S., Bourque, G., Liu, E. T., and Cheung, E. 2008. “Regulation of Estrogen Receptor-Mediated Long Range Transcription via Evolutionarily Conserved Distal Response Elements.” *Journal of Biological Chemistry* 283 (47): 32977–88. https://doi.org/10.1074/jbc.M802024200.

Papachristou, E. K., Kishore, K., Holding, A. N., Harvey, K., Roumeliotis, T. I., Chilamakuri, C. S. R., Omarjee, S., et al. 2018. “A Quantitative Mass Spectrometry-Based Approach to Monitor the Dynamics of Endogenous Chromatin-Associated Protein Complexes.” *Nature Communications* 9 (1): 2311. https://doi.org/10.1038/s41467-018-04619-5.

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., Davies, S., et al. 2009. “Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 27 (8): 1160–67. https://doi.org/10.1200/JCO.2008.18.1370.

Pastor, W. A., Aravind, L., and Rao, A. 2013. “TETonic Shift: Biological Roles of TET Proteins in DNA Demethylation and Transcription.” *Nature Reviews. Molecular Cell Biology* 14 (6): 341–56. https://doi.org/10.1038/nrm3589.

Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., et al. 2011. “Genome-Wide Mapping of 5-Hydroxymethylcytosine in Embryonic Stem Cells.” *Nature* 473 (7347): 394–97. https://doi.org/10.1038/nature10102.

Patnaik, M. M., Zahid, M. F., Lasho, T. L., Finke, C., Ketterling, R. L., Gangat, N., Robertson, K. D., Hanson, C. A., and Tefferi, A. 2016. “Number and type of TET2 mutations in chronic myelomonocytic leukemia and their clinical relevance” Blood Cancer J. 6 (9): e472. https://doi.org/10.1038/bcj.2016.82.

Paz, M. F., Avila, S., Fraga, M. F., Pollan, M., Capella, G., Peinado, M. A., Sanchez-Cespedes, M., Herman, J. G., and Esteller, M. 2002. “Germ-Line Variants in Methyl-Group Metabolism Genes and Susceptibility to DNA Methylation in Normal Tissues and Human Primary Tumors.” *Cancer Research* 62 (15): 4519–24. http://www.ncbi.nlm.nih.gov/pubmed/12154064.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., and Bejerano, G. 2013. “Enhancers: Five Essential Questions.” *Nature Reviews. Genetics* 14 (4): 288–95. https://doi.org/10.1038/nrg3458.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., et al. 2000. “Molecular Portraits of Human Breast Tumours.” *Nature* 406 (6797): 747–52. https://doi.org/10.1038/35021093.

Petterson, A., Chung, T. H., Tan, D., Sun, X., and Jia, X. Y. 2014. “RRHP: A Tag-Based Approach for 5-Hydroxymethylcytosine Mapping at Single-Site Resolution.” *Genome Biology* 15 (9): 456. https://doi.org/10.1186/s13059-014-0456-5.

Pickering, C. R., Zhou, J. H., Lee, J. J., Drummond, J. A., Peng, S. A., Saade, R. E., Tsai, K. Y., et al. 2014. “Mutational Landscape of Aggressive Cutaneous Squamous Cell Carcinoma.” *Clinical Cancer Research* 20 (24): 6582–92. https://doi.org/10.1158/1078-0432.CCR-14-1768.

Porter, W., Saville, B., Hoivik, D., and Safe, S. 1997. “Functional Synergy between the Transcription Factor Sp1 and the Estrogen Receptor.” *Molecular Endocrinology* 11 (11): 1569–80. https://doi.org/10.1210/mend.11.11.9916.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. 2001. “The P120 Catenin Partner Kaiso Is a DNA Methylation-Dependent Transcriptional Repressor.” *Genes & Development* 15 (13): 1613–18. https://doi.org/10.1101/gad.198501.

Putiri, E. L., Tiedemann, R. L., Thompson, J. J., Liu, C., Ho, T., Choi, J. H., and Robertson, K. D. 2014. “Distinct and Overlapping Control of 5-Methylcytosine and 5-Hydroxymethylcytosine by the TET Proteins in Human Cancer Cells.” *Genome Biology* 15 (6): R81. https://doi.org/10.1186/gb-2014-15-6-r81.

Qi, J., Zhang, X., Zhang, H. K., Yang, H. M., Zhou, Y. B., and Han, Z. G. 2006. “ZBTB34, a Novel Human BTB/POZ Zinc Finger Protein, Is a Potential Transcriptional Repressor.” *Molecular and Cellular Biochemistry* 290 (1–2): 159–67. https://doi.org/10.1007/s11010-006-9183-x.

Quinlan, A. R., and Hall, I. M. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Quivoron, C., Couronné, L, Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., et al. 2011. “TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis.” *Cancer Cell* 20 (1): 25–38. https://doi.org/10.1016/j.ccr.2011.06.003.

Rasmussen, K. D., Jia, G., Johansen, J. V., Pedersen, M. T., Rapin, N., Bagger, F. O., Porse, B. T., Bernard, O. A., Christensen, J., and Helin, K. 2015. “Loss of TET2 in Hematopoietic Cells Leads to DNA Hypermethylation of Active Enhancers and Induction of Leukemogenesis.” *Genes & Development* 29 (9): 910–22. https://doi.org/10.1101/gad.260174.115.

Rasmussen, K. D., and Helin, K. 2016. “Role of TET Enzymes in DNA Methylation, Development, and Cancer.” *Genes & Development* 30 (7): 733–50. https://doi.org/10.1101/gad.276568.115.

Rasmussen, K. D., Berest, I., Keβler, S., Nishimura, K., Simón-Carrasco, L., Vassiliou, G. S., Pedersen, M. T., Christensen, J., Zaugg, J. B., and Helin, K. 2019. “TET2 Binding to Enhancers Facilitates Transcription Factor Recruitment in Hematopoietic Cells.” *Genome Research* 29 (4): 564–75. https://doi.org/10.1101/gr.239277.118.

Ripperger, T., Gadzicki, D., Meindl, A., and Schlegelberger, B. 2009. “Breast Cancer Susceptibility: Current Knowledge and Implications for Genetic Counselling.” *European Journal of Human Genetics : EJHG* 17 (6): 722–31. https://doi.org/10.1038/ejhg.2008.212.

Robert, N. M., Tremblay, J. J., and Viger, R. S. 2002. “Friend of GATA (FOG)-1 and FOG-2 Differentially Repress the GATA-Dependent Activity of Multiple Gonadal Promoters.” *Endocrinology* 143 (10): 3963–73. https://doi.org/10.1210/en.2002-220280.

Robertson, J. F. R. 2007. “Fulvestrant (Faslodex®) How to Make a Good Drug Better.” *The Oncologist* 12 (7): 774–84. https://doi.org/10.1634/theoncologist.12-7-774.

Rollins, D. A., Coppo, M., and Rogatsky, I. 2015. “Minireview: Nuclear Receptor Coregulators of the P160 Family: Insights into Inflammation and Metabolism.” *Molecular Endocrinology (Baltimore, Md.)* 29 (4): 502–17. https://doi.org/10.1210/me.2015-1005.

Rønneberg, J. A., Fleischer, T., Solvang, H. K., Nordgard, S. H., Edvardsen, H., Potapenko, I., Nebdal, D., et al. 2011. “Methylation Profiling with a Panel of Cancer Related Genes: Association with Estrogen Receptor, TP53 Mutation Status and Expression Subtypes in Sporadic Breast Cancer.” *Molecular Oncology* 5 (1): 61–76. https://doi.org/10.1016/j.molonc.2010.11.004.

Ross-Innes, C. S., Stark, R., Holmes, K. A., Schmidt, D., Spyrou, C., Russell, R., Massie, C. E., Vowler, S. L., Eldridge, M., and Carroll, J. S. 2010. “Cooperative Interaction between Retinoic Acid Receptor-Alpha and Estrogen Receptor in Breast Cancer.” *Genes & Development* 24 (2): 171–82. https://doi.org/10.1101/gad.552910.

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., et al. 2012. “Differential Oestrogen Receptor Binding Is Associated with Clinical Outcome in Breast Cancer.” *Nature* 481 (7381): 389–93. https://doi.org/10.1038/nature10730.

Rueda, O. M., Sammut, S. J., Seoane, J. A., Chin, S. F., Caswell-Jin, J. L., Callari, M., Batra, R., et al. 2019. “Dynamics of Breast-Cancer Relapse Reveal Late-Recurring ER-Positive Genomic Subgroups.” *Nature* 567 (7748): 399–404. https://doi.org/10.1038/s41586-019-1007-8.

Safe, S. 2001. “Transcriptional Activation of Genes by 17 Beta-Estradiol through Estrogen Receptor-Sp1 Interactions.” *Vitamins and Hormones* 62: 231–52. http://www.ncbi.nlm.nih.gov/pubmed/11345900.

Schulz, W. A., Steinhoff, C., and Florl, A. R. 2006. “Methylation of Endogenous Human Retroelements in Health and Disease.” In *DNA Methylation: Development, Genetic Disease and Cancer*, 310:211–50. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-31181-5_11.

Schwabe, J. W., Chapman, L., and Rhodes, D. 1995. “The Oestrogen Receptor Recognizes an Imperfectly Palindromic Response Element through an Alternative Side-Chain Conformation.” *Structure (London, England : 1993)* 3 (2): 201–13. http://www.ncbi.nlm.nih.gov/pubmed/7735836.

Scourzic, L., Mouly, E., and Bernard, O. A. 2015. “TET Proteins and the Control of Cytosine Demethylation in Cancer.” *Genome Medicine* 7 (1): 9. https://doi.org/10.1186/s13073-015-0134-6.

Sekiya, T., and Zaret, K. S. 2007. “Repression by Groucho/TLE/Grg Proteins: Genomic Site Recruitment Generates Compacted Chromatin In Vitro and Impairs Activator Binding In Vivo.” *Molecular Cell* 28 (2): 291–303. https://doi.org/10.1016/j.molcel.2007.10.002.

Serandour, A. A., Brown, G. D., Cohen, J. D., and Carroll, J. S. 2013. “Development of an Illumina-Based ChIP-Exonuclease Method Provides Insight into FoxA1-DNA Binding Properties.” *Genome Biology* 14 (12): R147. https://doi.org/10.1186/gb-2013-14-12-r147.

Sflomos, G., Dormoy, V., Metsalu, T., Jeitziner, R., Battista, L., Scabia, V., Raffoul, W., et al. 2016. “A Preclinical Model for ERα-Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response.” *Cancer Cell* 29 (3): 407–22. https://doi.org/10.1016/j.ccell.2016.02.002.

Shang, Y., and Brown, M. 2002. “Molecular Determinants for the Tissue Specificity of SERMs.” *Science* 295 (5564): 2465–68. https://doi.org/10.1126/science.1068537.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A., and Brown, M. 2000. “Cofactor Dynamics and Sufficiency in Estrogen Receptor-Regulated Transcription.” *Cell* 103 (6): 843–52. https://doi.org/10.1016/s0092-8674(00)00188-4.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A. C., Fung, H. L., Zhang, K., and Zhang, Y. 2013. “Genome-Wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics.” *Cell* 153 (3): 692–706. https://doi.org/10.1016/j.cell.2013.04.002.

Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., and Greene, G. L. 1998. “The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen.” *Cell* 95 (7): 927–37. https://doi.org/10.1016/s0092-8674(00)81717-1.

Shlyueva, D., Stampfel, G., and Stark, A. 2014. “Transcriptional Enhancers: From Properties to Genome-Wide Predictions.” *Nature Reviews Genetics* 15 (4): 272–86. https://doi.org/10.1038/nrg3682.

Shoker, B. S., Jarvis, C., Clarke, R. B., Anderson, E., Hewlett, J., Davies, M. P., Sibson, D. R., and Sloane, J. P. 1999. “Estrogen Receptor-Positive Proliferating Cells in the Normal and Precancerous Breast.” *The American Journal of Pathology* 155 (6): 1811–15. https://doi.org/10.1016/S0002-9440(10)65498-3.

Simpson, E.R. 2003. “Sources of Estrogen and Their Importance.” *The Journal of Steroid Biochemistry and Molecular Biology* 86 (3–5): 225–30. https://doi.org/10.1016/S0960-0760(03)00360-1.

Song, C. X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S. Q., Lin, L., Street, C., et al. 2013. “Genome-Wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming.” *Cell* 153 (3): 678–91. https://doi.org/10.1016/j.cell.2013.04.001.

Song, J., and Pfeifer, G. P. 2016. “Are There Specific Readers of Oxidized 5-Methylcytosine Bases?” *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 38 (10): 1038–47. https://doi.org/10.1002/bies.201600126.

Song, Y., van den Berg, P. R., Markoulaki, S., Soldner, F., Dall’Agnese, A., Henninger, J,. E., Drotar, J., et al. 2019. “Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs.” *Molecular Cell*, August. https://doi.org/10.1016/J.MOLCEL.2019.06.045.

Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., et al. 2001. “Gene Expression Patterns of Breast Carcinomas Distinguish Tumor Subclasses with Clinical Implications.” *Proceedings of the National Academy of Sciences* 98 (19): 10869–74. https://doi.org/10.1073/pnas.191367098.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., et al. 2003. “Repeated Observation of Breast Tumor Subtypes in Independent Gene Data Sets.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (14): 8418. https://doi.org/10.1073/PNAS.0932692100.

Speirs, V., Skliris, G. P., Burdall, S. E., and Carder, P. J. 2002. “Distinct Expression Patterns of ER Alpha and ER Beta in Normal Human Mammary Gland.” *Journal of Clinical Pathology* 55 (5): 371–74. https://doi.org/10.1136/jcp.55.5.371.

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., et al. 2013. “Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives.” *Cell* 152 (5): 1146–59. https://doi.org/10.1016/J.CELL.2013.02.004.

Spruijt, C. G., and Vermeulen, M. 2014. “DNA Methylation: Old Dog, New Tricks?” *Nature Publishing Group* 21. https://doi.org/10.1038/nsmb.2910.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., et al. 2011. “DNA-Binding Factors Shape the Mouse Methylome at Distal Regulatory Regions.” *Nature* 480 (7378): 490–95. https://doi.org/10.1038/nature10716.

Stark, R., and Brown, G. 2013. “DiffBind : Differential Binding Analysis of ChIP-Seq Peak Data.” https://bioconductor.statistik.tu-dortmund.de/packages/2.13/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

Stender, J. D., Kim, K., Charn, T. H., Komm, B., Chang, K. C. N., Kraus, W. L., Benner, C., Glass, C. K., and Katzenellenbogen, B. S. 2010. “Genome-Wide Analysis of Estrogen Receptor Alpha DNA Binding and Tethering Mechanisms Identifies Runx1 as a Novel Tethering Factor in Receptor-Mediated Transcriptional Activation.” *Molecular and Cellular Biology* 30 (16): 3943–55. https://doi.org/10.1128/MCB.00118-10.

Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., et al. 2012. “The Landscape of Cancer Genes and Mutational Processes in Breast Cancer.” *Nature* 486 (7403): 400–404. https://doi.org/10.1038/nature11017.

Stern, J. L., Paucek, R. D., Huang, F. W., Ghandi, M., Nwumeh, R., Costello, J. C., and Cech, T. R. 2017. “Allele-Specific DNA Methylation and Its Interplay with Repressive Histone Marks at Promoter-Mutant TERT Genes.” *Cell Reports* 21 (13): 3700–3707. https://doi.org/10.1016/j.celrep.2017.12.001.

Stone, A., Valdés-Mora, F., Gee, J. M. W., Farrow, L., McClelland, R. A., Fiegl, H., Dutkowski, C., et al. 2012. “Tamoxifen-Induced Epigenetic Silencing of Oestrogen-Regulated Genes in Anti-Hormone Resistant Breast Cancer.” Edited by Bernard W. Futscher. *PLoS ONE* 7 (7): e40466. https://doi.org/10.1371/journal.pone.0040466.

Stossi, F., Likhite, V. S., Katzenellenbogen, J. A., and Katzenellenbogen, B. S. 2006. “Estrogen-Occupied Estrogen Receptor Represses Cyclin G2 Gene Expression and Recruits a Repressor Complex at the Cyclin G2 Promoter.” *The Journal of Biological Chemistry* 281 (24): 16272–78. https://doi.org/10.1074/jbc.M513405200.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S. E. 2011. “5-Hydroxymethylcytosine Is Associated with Enhancers and Gene Bodies in Human Embryonic Stem Cells.” *Genome Biology* 12 (6): R54. https://doi.org/10.1186/gb-2011-12-6-r54.

Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., and Leonhardt, H. 2010. “Sensitive Enzymatic Quantification of 5-Hydroxymethylcytosine in Genomic DNA.” *Nucleic Acids Research* 38 (19): e181–e181. https://doi.org/10.1093/nar/gkq684.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., et al. 2009. “Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1.” *Science (New York, N.Y.)* 324 (5929): 930–35. https://doi.org/10.1126/science.1170116.

Takaku, M., Grimm, S. A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H. G., Archer, T. K., Machida, S., Kurumizaka, H., and Wade, P. A. 2016. “GATA3-Dependent Cellular Reprogramming Requires Activation-Domain Dependent Recruitment of a Chromatin Remodeler.” *Genome Biology* 17 (1): 36. https://doi.org/10.1186/s13059-016-0897-0.

Takaku, M., Grimm, S. A., Roberts, J. D., Chrysovergis, K., Bennett, B. D., Myers, P., Perera, L., Tucker, C. J., Perou, C. M., and Wade, P. A. 2018. “GATA3 Zinc Finger 2 Mutations Reprogram the Breast Cancer Transcriptional Network.” *Nature Communications* 9. https://doi.org/10.1038/S41467-018-03478-4.

Takayama, K., Misawa, A., Suzuki, T., Takagi, K., Hayashizaki, Y., Fujimura, T., Homma, Y., Takahashi, S., Urano, T., and Inoue, S. 2015. “TET2 Repression by Androgen Hormone Regulates Global Hydroxymethylation Status and Prostate Cancer Progression.” *Nature Communications* 6 (1): 8219. https://doi.org/10.1038/ncomms9219.

Tan, S. K., Lin, Z. H., Chang, C. W., Varang, V., Chng, K. R., Pan, Y. F., Yong, E. L., Sung, W. K., Sung, W. K, and Cheung, E. 2011. “AP-2γ Regulates Oestrogen Receptor-Mediated Long-Range Chromatin Interaction and Gene Transcription.” *The EMBO Journal* 30 (13): 2569–81. https://doi.org/10.1038/emboj.2011.151.

Taylor, S. E. B, Li, Y. H., Smeriglio, P., Rath, M., Wong, W. H., and Bhutani, N. 2016. “Stable 5-Hydroxymethylcytosine (5hmC) Acquisition Marks Gene Activation During Chondrogenic Differentiation.” *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research* 31 (3): 524–34. https://doi.org/10.1002/jbmr.2711.

Theodorou, V., Stark, R., Menon, S., and Carroll, J. S. 2013. “GATA3 Acts Upstream of FOXA1 in Mediating ESR1 Binding by Shaping Enhancer Accessibility.” *Genome Research* 23 (1): 12–22. https://doi.org/10.1101/gr.139469.112.

Tikkanen, M. K., Carter, D. J., Harris, A. M., Le, H. M., Azorsa, D. O., Meltzer, P. S., and Murdoch, F. E. 2000. “Endogenously Expressed Estrogen Receptor and Coactivator AIB1 Interact in MCF-7 Human Breast Cancer Cells.” *Proceedings of the National Academy of Sciences* 97 (23): 12536–40. https://doi.org/10.1073/pnas.220427297.

Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerød, A., et al. 2004. “Mutation of GATA3 in Human Breast Tumors.” *Oncogene* 23 (46): 7669–78. https://doi.org/10.1038/sj.onc.1207966.

Valinluck, V., and Sowers, L. C. 2007. “Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1.” *Cancer Research* 67 (3): 946–50. https://doi.org/10.1158/0008-5472.CAN-06-3123.

Varlakhanova, N., Snyder, C., Jose, S., Hahm, J. B., and Privalsky, M. L. 2010. “Estrogen Receptors Recruit SMRT and N-CoR Corepressors through Newly Recognized Contacts between the Corepressor N Terminus and the Receptor DNA Binding Domain.” *Molecular and Cellular Biology* 30 (6): 1434–45. https://doi.org/10.1128/MCB.01002-09.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., et al. 2013. “Tet Proteins Connect the O-Linked N-Acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells.” *Molecular Cell* 49 (4): 645–56. https://doi.org/10.1016/J.MOLCEL.2012.12.019.

Wagner, S., Weber, S., Kleinschmidt, M. A., Nagata, K., and Bauer, U. M. 2006. “SET-Mediated Promoter Hypoacetylation Is a Prerequisite for Coactivation of the Estrogen-Responsive PS2 Gene by PRMT1.” *Journal of Biological Chemistry* 281 (37): 27242–50. https://doi.org/10.1074/JBC.M605172200.

Wakeling, A. E., Dukes, M., and Bowler, J. 1991. “A Potent Specific Pure Antiestrogen with Clinical Potential.” *Cancer Research* 51 (15): 3867–73. http://www.ncbi.nlm.nih.gov/pubmed/1855205.

Wang, G. G., Allis, C. D., and Chi, P. 2007. “Chromatin Remodeling and Cancer, Part II: ATP-Dependent Chromatin Remodeling.” *Trends in Molecular Medicine* 13 (9): 373. https://doi.org/10.1016/J.MOLMED.2007.07.004.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., et al. 2015. “WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation.” *Molecular Cell* 57 (4): 662–73. https://doi.org/10.1016/j.molcel.2014.12.023.

Wang, L., Ozark, P. A., Smith, E. R., Zhao, Z., Marshall, S. A., Rendleman, E. J., Piunti, A., et al. 2018. “TET2 Coactivates Gene Expression through Demethylation of Enhancers.” *Sci. Adv*. Vol. 4. http://advances.sciencemag.org/.

Watson, P. J., Fairall, L., and Schwabe, J. W. R. 2012. “Nuclear Hormone Receptor Co-Repressors: Structure and Function.” *Molecular and Cellular Endocrinology* 348 (2): 440–49. https://doi.org/10.1016/j.mce.2011.08.033.

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., and Helin, K. 2011. “TET1 and Hydroxymethylcytosine in Transcription and DNA Methylation Fidelity.” *Nature* 473 (7347): 343–48. https://doi.org/10.1038/nature10066.

Wilson, A. S., Power, B. E., and Molloy, P. L. 2007. “DNA Hypomethylation and Human Diseases.” *Biochimica et Biophysica Acta - Reviews on Cancer* 1775 (1): 138–62. https://doi.org/10.1016/j.bbcan.2006.08.007.

Wu, H., D’Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. 2011. “Genome-Wide Analysis of 5-Hydroxymethylcytosine Distribution Reveals Its Dual Function in Transcriptional Regulation in Mouse Embryonic Stem Cells.” *Genes & Development* 25 (7): 679–84. https://doi.org/10.1101/GAD.2036011.

Wu, H., and Zhang, Y. 2011. “Tet1 and 5-Hydroxymethylation.” *Cell Cycle* 10 (15): 2428–36. https://doi.org/10.4161/cc.10.15.16930.

Wu, X. and Zhang, Y. 2017. “TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond.” *Nature Reviews Genetics* 18 (9): 517–34. https://doi.org/10.1038/nrg.2017.33.

Xu, C., Liu, K. Lei, M., Yang, A., Li, Y., Hughes, T. R., and Min, J. 2018. “DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants.” *Structure* 26 (1): 85-95.e3. https://doi.org/10.1016/j.str.2017.11.022.

Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T. R., and Min, J. 2018. “DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants Highlights d Solved Structures of Human CXXC/CXXC-PHD Domains with Respective DNA Ligands d Human CXXC Domains Display Different DNA-Binding Selectivities d Human CXXC Domain.” https://doi.org/10.1016/j.str.2017.11.022.

Yang, H., Liu, Y., Bai, F., Zhang, J. Y., Ma, S. H., Liu, J., Xu, Z. D., et al. 2013. “Tumor Development Is Associated with Decrease of TET Gene Expression and 5-Methylcytosine Hydroxylation.” *Oncogene* 32 (5): 663–69. https://doi.org/10.1038/onc.2012.67.

Yang, L., Yu, S. J., Hong, Q., Yang, Y., and Shao, Z. M. 2015. “Reduced Expression of TET1, TET2, TET3 and TDG MRNAs Are Associated with Poor Prognosis of Patients with Early Breast Cancer.” Edited by Xin-Yuan Guan. *PLOS ONE* 10 (7): e0133896. https://doi.org/10.1371/journal.pone.0133896.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., et al. 2017. “Impact of Cytosine Methylation on DNA Binding Specificities of Human Transcription Factors.” *Science* 356 (6337): eaaj2239. https://doi.org/10.1126/science.aaj2239.

Yoon, G. H., Chan, D. W., Reynolds, A. B., Qin, J., and Wong, J. 2003. “N-CoR Mediates DNA Methylation-Dependent Repression through a Methyl CpG Binding Protein Kaiso.” *Molecular Cell* 12 (3): 723–34. http://www.ncbi.nlm.nih.gov/pubmed/14527417.

Zaret, K. S., and Carroll, J. S. 2011. “Pioneer Transcription Factors: Establishing Competence for Gene Expression.” *Genes & Development* 25 (21): 2227–41. https://doi.org/10.1101/gad.176826.111.

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu,Y., Li, X., Zhao, D., et al. 2015. “Tet2 Is Required to Resolve Inflammation by Recruiting Hdac2 to Specifically Repress IL-6.” *Nature* 525 (7569): 389–93. https://doi.org/10.1038/nature15252.

Zhang, Y. W., Wang, Z., Xie, W., Cai, Y., Xia, L., Easwaran, H., Luo, J., Yen, R. W. C., Li, Y., and Baylin, S. B. 2017. “Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress.” *Molecular Cell* 65 (2): 323–35. https://doi.org/10.1016/j.molcel.2016.12.013.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., et al. 2008. “Model-Based Analysis of ChIP-Seq (MACS).” *Genome Biology* 9 (9): R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhong, J., Li, X., Cai, W., Wang, Y., Dong, S., Yang, J., Zhang, J., et al. 2017. “TET1 Modulates H4K16 Acetylation by Controlling Auto-Acetylation of HMOF to Affect Gene Regulation and DNA Repair Function.” *Nucleic Acids Research* 45 (2): 672–84. https://doi.org/10.1093/nar/gkw919.

Zhou, T., Xiong, J., Wang, W., Yang, N., Wong, J., Zhu, B., and Xu, R. M. 2014. “Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2.” *Molecular Cell* 54 (5): 879–86. https://doi.org/10.1016/j.molcel.2014.04.003.

Zhou, V. W., Goren, A., and Bernstein, B. E. 2011. “Charting Histone Modifications and the Functional Organization of Mammalian Genomes.” *Nature Reviews Genetics* 12 (1): 7–18. https://doi.org/10.1038/nrg2905.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., De Jager, P. L., et al. 2013. “Charting a Dynamic DNA Methylation Landscape of the Human Genome.” *Nature* 500 (7463): 477. https://doi.org/10.1038/NATURE12433.

Zubairy, S., and Oesterreich, S. 2013. “Estrogen-repressed genes – key mediators of estrogen action?” *Breast Cancer Res.* 7 (4): 163–164 https://doi.org/10.1186/bcr1271.

Zwart, W., de Leeuw, R., Rondaij, M., Neefjes, J., Mancini, M. A., and Michalides, R. 2010. “The Hinge Region of the Human Estrogen Receptor Determines Functional Synergy between AF-1 and AF-2 in the Quantitative Response to Estradiol and Tamoxifen.” *Journal of Cell Science* 123 (8): 1253–61. https://doi.org/10.1242/jcs.061135.